

**AUTOMATED TAILORING OF CLINICAL
PERFORMANCE FEEDBACK IN LOW-RESOURCE
SETTINGS**

by

Zachary J Landis Lewis

Master of Science in Biomedical Informatics, University of
Pittsburgh, 2010

Submitted to the Graduate Faculty of
certain Department of Biomedical Informatics in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH
SCHOOL OF MEDICINE

This dissertation was presented

by

Zachary J Landis Lewis

It was defended on

August 28th 2014

and approved by

Rebecca S Jacobson, MD, MS, Professor, Department of Biomedical Informatics,

University of Pittsburgh

Gerald P Douglas, PhD, Assistant Professor, Department of Biomedical Informatics,

University of Pittsburgh

Harry Hochheiser, PhD, Assistant Professor, Department of Biomedical Informatics,

University of Pittsburgh

Matthew Kam, PhD, Google

Dissertation Director: Rebecca S Jacobson, MD, MS, Professor, Department of Biomedical
Informatics, University of Pittsburgh

AUTOMATED TAILORING OF CLINICAL PERFORMANCE FEEDBACK IN LOW-RESOURCE SETTINGS

Zachary J Landis Lewis, PhD

University of Pittsburgh, 2014

A patient-centered, continuously learning healthcare system is a compelling vision for the future of healthcare, introduced by the Institute of Medicine. A key part of this vision is the creation of feedback loops to support continuous clinical learning and behavior change. Opportunities to generate clinical performance feedback are increasing, due to globally unprecedented growth in the adoption of eHealth. These opportunities are especially promising in low-income countries where a critical problem is poor performance of healthcare providers that lowers the quality of care.

Clinical audit and feedback, defined as the provision of performance summaries to healthcare providers, teams, and organizations, is widely used for quality improvement and the implementation of evidence-based practice. Evidence shows that clinical audit and feedback can significantly improve compliance with desired practice, but it is unclear when and how it is most effective. Psychological theories offer rigorously evaluated theoretical causal mechanisms that may explain when feedback is likely to be effective for clinical learning and behavior change, but these have rarely been used to inform the design of feedback interventions. In addition to uncertainty regarding the effect of feedback on clinical performance, a critical challenge for using eHealth data to automate the delivery of feedback is understanding data quality for the purpose of performance measurement. To overcome the dual challenges of variable data quality and performance feedback effectiveness, I propose a novel, theory-informed approach for generating clinical performance feedback: automated feedback message tailoring.

This research explores evidence, theories, methods, and clinical settings that establish a foundation of knowledge for the automated tailoring of feedback messages. I developed and applied this knowledge within antiretroviral therapy clinics in Malawi, Africa, where an electronic medical record system is routinely used, to understand the potential impact of feedback message tailoring in low-resource settings. This work introduces a novel information tool that may enable clinical supervisors to use existing eHealth data to provide more effective performance feedback, and which may support the testing of hypotheses about the effect of tailored feedback messages on clinical performance.

TABLE OF CONTENTS

1.0	INTRODUCTION	1
1.1	BACKGROUND	1
1.2	DESCRIPTION OF THE PROBLEM	2
1.3	SIGNIFICANCE OF THIS RESEARCH	4
2.0	E-HEALTH IN LOW-INCOME COUNTRIES	5
2.1	DATA QUALITY	7
3.0	SETTING: MALAWI	9
3.1	TREATMENT OF HIV/AIDS	9
3.2	NATIONAL ART EMR	10
4.0	IMPLEMENTATION SCIENCE	12
4.1	THE KNOWLEDGE TO ACTION CYCLE	13
4.2	CONSOLIDATED FRAMEWORK FOR IMPLEMENTATION RESEARCH	15
4.3	PEPFAR Implementation Science Framework	19
5.0	IMPROVING HEALTHCARE IN LOW-RESOURCE SETTINGS	21
5.1	TASK-SHIFTING	21
5.2	CLINICAL SUPERVISION AND MENTORING	22
5.3	CLINICAL PRACTICE GUIDELINES	24
5.3.1	Guideline development	24
5.3.2	Guideline implementation	26
5.4	CONCLUSION	32
6.0	AUDIT AND FEEDBACK	33
6.1	EXAMPLES OF AUDIT AND FEEDBACK INTERVENTIONS	37

6.1.1	Audit and feedback targeting antimicrobial stewardship	38
6.1.2	Process-focused AF: Guideline-adherence AF in Lao PDR	40
6.1.3	Outcome-focused AF: Critical incident AF for maternal mortality in Malawi	41
6.2	EFFECTIVENESS OF AUDIT AND FEEDBACK	42
6.3	EFFECTIVENESS OF AUDIT AND FEEDBACK IN LOW AND MIDDLE- INCOME COUNTRIES	48
6.4	CONCLUSION	51
7.0	USING THEORY TO INFORM IMPLEMENTATION	52
7.1	THE THEORETICAL DOMAINS FRAMEWORK	52
7.2	COM-B FRAMEWORK	53
7.2.1	Capability barriers	54
7.2.2	Opportunity barriers	55
7.2.3	Motivational barriers	56
7.3	MENU OF CONSTRUCTS	58
7.4	SUMMARY	59
8.0	FEEDBACK THEORY	60
8.1	FEEDBACK-STANDARD COMPARISON THEORIES	61
8.2	FEEDBACK PROCESS MODEL	64
8.3	FEEDBACK INTERVENTION THEORY	68
8.4	REGULATORY FOCUS THEORY	73
8.5	GROUP FEEDBACK PROCESSING	78
8.6	SUMMARY	81
9.0	TAILORING CLINICAL PERFORMANCE FEEDBACK	82
9.1	FEEDBACK TAILORING IN MEDICAL KNOWLEDGE-BASED SYS- TEMS	82
9.1.1	Ontologies and knowledge representation artifacts	83
9.1.2	Probabilistic knowledge representations	85
9.1.3	Tailoring of clinical feedback	85
9.2	COMPUTER-TAILORED HEALTH COMMUNICATION	86

9.2.1	AUTOMATED TAILORING OF HEALTH MESSAGES	88
9.3	SUMMARY	90
10.0	RESEARCH DESIGN AND METHODS	92
11.0	MEASURING INDIVIDUAL CLINICAL PERFORMANCE USING E-HEALTH DATA IN LOW-RESOURCE SETTINGS	94
11.1	METHODS	94
11.2	RESULTS	98
11.3	DISCUSSION	103
11.4	LIMITATIONS	104
11.5	CONCLUSION	105
12.0	BARRIERS TO USING E-HEALTH DATA FOR PERFORMANCE FEEDBACK IN LOW-RESOURCE SETTINGS	106
12.1	METHODS	106
12.1.1	Setting	106
12.1.2	Data collection	106
12.1.2.1	Interviews:	107
12.1.2.2	Observations:	109
12.1.2.3	Informant feedback meetings:	109
12.1.3	Data analysis:	109
12.2	QUALITATIVE FINDINGS	110
12.2.1	Provider rotations	110
12.2.2	Disruptions to care processes	111
12.2.3	User acceptance of eHealth	112
12.2.4	Performance indicator lifespan	113
12.3	DISCUSSION	115
12.4	LIMITATIONS	118
12.5	CONCLUSION	118
13.0	DESIGNING KNOWLEDGE-BASED FEEDBACK TAILORING . .	119
13.1	METHODS	121
13.2	RESULTS	122

13.2.1	Low-performance of unnecessary antibiotic prescribing for acute respiratory infection	126
13.2.2	Automated feedback message tailoring	130
13.2.3	Using performance features to indicate construct salience	131
13.2.4	System architecture	134
13.2.4.1	Tailoring knowledge base	136
13.2.4.2	eHealth database	140
13.2.4.3	Performance measurement process	140
13.2.4.4	Performance database	141
13.2.4.5	Message tailoring process	141
13.2.4.6	Identify performance features	142
13.2.4.7	Infer construct salience	142
13.2.4.8	Assess component relevance	142
13.2.4.9	Prioritize messages	143
13.3	DISCUSSION	144
13.3.1	LIMITATIONS	145
13.4	CONCLUSION	146
14.0	AUTOMATED FEEDBACK TAILORING IN LOW-RESOURCE SETTINGS	147
14.1	METHODS	147
14.1.1	Performance measurement	148
14.1.2	Message tailoring and prioritization	150
14.1.3	Potential impact of tailoring	151
14.1.3.1	Performance gaps	151
14.1.3.2	Message variability	152
14.2	RESULTS	153
14.2.1	Performance measurement	153
14.2.2	Feedback message tailoring	160
14.2.3	Potential impact of tailoring	160
14.2.3.1	Performance gaps	160

14.2.3.2 Message variability	167
14.3 DISCUSSION	177
14.4 LIMITATIONS	179
14.5 CONCLUSION	180
15.0 DISCUSSION	181
15.1 LIMITATIONS	183
15.2 FUTURE WORK	183
16.0 CONCLUSIONS	186
APPENDIX. INTERVIEW GUIDE	187
BIBLIOGRAPHY	190

LIST OF TABLES

1	Auditable recommendations and proof-of-concept adherence feedback.	101
2	Characteristics of interview participants (N = 32)	108
3	Mapping antibiotic prescribing barriers to theoretical constructs.	124
4	Individual antibiotic prescribing performance data with a peer benchmark . .	127
5	Definitions of tailoring knowledge-base components.	137
6	Message tailoring process from a prototype feedback tailoring system.	141
7	Four performance indicators mapped to Malawi's national ART guideline. . .	149
8	Totals of message tailoring rules used in the message tailoring process.	151
9	Individual monthly report totals summary statistics.	153
10	Mean performance of 11 ART clinics in Malawi, 10/2011 to 9/2013.	154
11	Mean performance of 11 ART clinics in Malawi for four indicators.	155
12	Characteristics of data collected and analyzed for feedback message tailoring.	160
13	Percentage of months containing performance gaps at 11 ART clinics in Malawi.	161
14	Mean monthly absolute differences in percentage of prioritized message types.	172

LIST OF FIGURES

1	Interaction between performance measurement and feedback effectiveness. . .	3
2	The knowledge to action (KTA) cycle (Graham et al. 2006).	14
3	Consolidated Framework for Implementation Research (Damschroder et al. 2009).	16
4	Prototype feedback report for diabetes care used by Ivers et al. 2010.	34
5	The Plan-Do-Study-Act cycle.	35
6	A model of actionable feedback by Hysong, Best and Pugh, 2006.	45
7	The COM-B framework for understanding behavior (Michie et al. 2011). . . .	54
8	A cognitive model of feedback processing from Ilgen, Taylor and Fisher, 1979.	65
9	Regulatory Focus and Feedback Sign	74
10	Methods for identification of auditable guideline recommendations.	95
11	Identification and assessment of recommendations.	100
12	Distribution of average monthly encounters for recommendations in Table 1. .	102
13	A menu of tailored messages for a low-performaning healthcare provider. . . .	129
14	System architecture for an automated AF message tailoring system.	135
15	Monthly pediatric height recording performance, 10/2011 to 9/2013.	156
16	Monthly weight recording performance, 10/2011 to 9/2013.	157
17	Monthly WHO clinical staging performance, 10/2011 to 9/2013.	158
18	Monthly CPT prescribing performance, 10/2011 to 9/2013.	159
19	Mean monthly performance gap totals, 10/2011 to 9/2013.	162
20	Performance gap totals for pediatric height recording, 10/2011 to 9/2013. . .	163
21	Performance gap totals for weight recording, 10/2011 to 9/2013.	164

22	Performance gap totals for WHO clinical staging, 10/2011 to 9/2013.	165
23	Performance gap totals for CPT prescribing, 10/2011 to 9/2013.	166
24	Mean percentages of prioritized message types.	167
25	Mean percentage of tailored messages prioritized for pediatric height recording.	168
26	Mean percentage of tailored messages prioritized for patient weight recording.	169
27	Mean percentage of tailored messages prioritized for WHO clinical staging. .	170
28	Mean percentage of tailored messages prioritized for CPT prescribing. . . .	171
29	Mean absolute percentage change in messages for pediatric height recording. .	173
30	Mean absolute percentage change in messages for patient weight recording. .	174
31	Mean absolute percentage change in messages for WHO clinical staging. . . .	175
32	Mean absolute percentage change in messages for CPT prescribing.	176

1.0 INTRODUCTION

1.1 BACKGROUND

Globally there are significant gaps between best practices drawn from medical evidence and decisions made by healthcare professionals.¹ These gaps exist both in high-income countries² and in low- and middle-income countries.³ Closing these gaps is increasingly difficult because of accelerating rates of the production of biomedical knowledge and the increasing complexity of healthcare systems.⁴

A globally unprecedented change in healthcare systems that coincides with increasing knowledge production and healthcare complexity is the adoption of electronic health information technology (eHealth).⁵⁻⁷ The routine use of eHealth has motivated a surge of interest in pairing large quantities of biomedical data with techniques to analyze and draw meaning from them, also known as big data.⁸ The Institute of Medicine has established a vision for addressing these challenges and opportunities in the creation of the continuously learning health system.⁴ A central component of this vision is the use of digital infrastructure to develop feedback loops that support clinical learning and behavior change to improve the delivery of patient-centered care.

Audit and feedback (AF) is defined as the provision of clinical performance summaries to healthcare providers, teams, and organizations.⁹ AF includes a heterogeneous set of approaches commonly used within multi-faceted interventions to support learning and behavior change for healthcare quality improvement. Evidence from the most recent Cochrane review, based on 140 clinical trials of AF, shows that AF can significantly improve compliance with desired practice, but that it is unclear which approaches, under which circumstances, will work.¹⁰ Given the relatively limited insights produced by AF trials to date, AF researchers

have called for a shift towards comparative effectiveness studies, evaluating how and when AF intervention components will work, rather than its overall effectiveness.¹¹ Furthermore, AF researchers have argued for the explicit use of psychological theory to identify and test theoretical causal mechanisms that AF operates on to change behavior support clinical learning.^{9,12}

I believe that big data is creating unprecedented opportunities to understand how and when AF can be used to sustain continuously learning healthcare systems, and that the most promising approach to understanding AF in clinical settings is via the explicit use of psychological theory. However, a key challenge for this work is the issue of clinical data quality for the purpose of performance measurement. Our understanding of how to improve and maintain clinical data quality is increasing^{13,14} but poor data quality persists as an important limitation for big data. Thus, two types of uncertainty must be managed to successfully automate audit and feedback using clinical data: the effectiveness of performance feedback, and the degree to which clinical data are fit for performance measurement.

1.2 DESCRIPTION OF THE PROBLEM

The purpose of this research is firstly to understand the interacting challenges of clinical performance measurement and feedback effectiveness in low-resource settings (Figure 1), and secondly to design and formatively evaluate a novel information system that may overcome these challenges. As more clinical data becomes available for analysis, more opportunities are created to measure performance and to provide feedback. However, for each of these opportunities, the quality of the data and the likely effect of the feedback on performance must be determined to understand when feedback is relevant to individual healthcare providers' goals of improving the quality of care.

To understand challenges of performance measurement associated with data quality, we recognize that data quality is a function of the degree to which data contain errors, and is fundamentally determined by the expectations of the data consumer for a specific purpose of data use.^{14,15} Therefore, the quality of data for the purpose of performance measurement

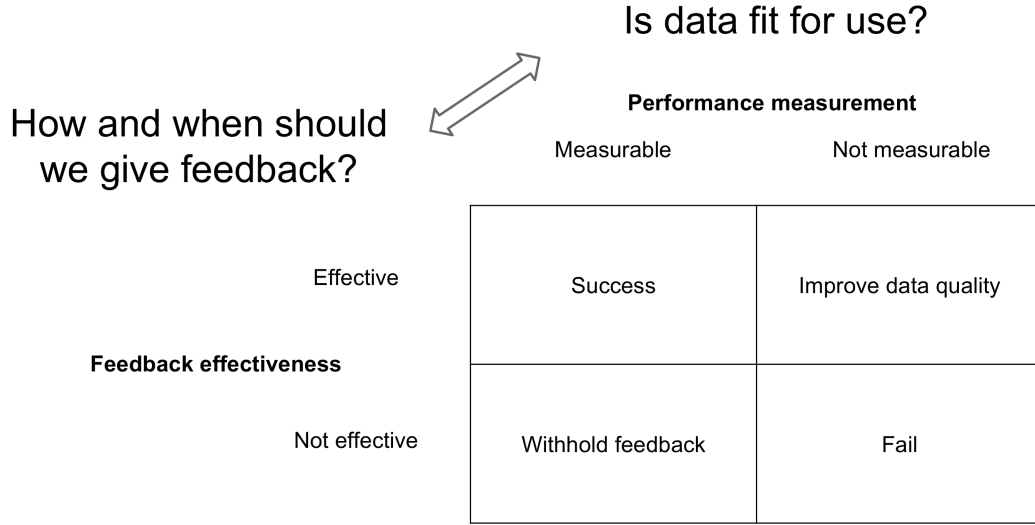


Figure 1: Interaction between performance measurement and feedback effectiveness.

can be expected to vary with sources of error introduced into the data, and according to variable expectations of individuals for a specific behavior that is being measured within a specific clinical context.

To understand feedback effectiveness, we can reason that AF interventions will be more effective when their components influence barriers to behavior change.^{16,17} However, barriers to behavior change differ across individual healthcare providers, stemming from differences in providers' training, knowledge, work experience, personality and other individual characteristics.¹⁸ Furthermore, barriers to change may be dynamic, as providers' beliefs, motivations, and perceptions are influenced by ongoing changes in the healthcare organization, the complexity of which is widely recognized.¹⁹ Therefore, performance feedback that is tailored for healthcare providers' individual and situational barriers to behavior change is more likely to contribute to improved performance.

To overcome the dual challenges of variable data quality and performance feedback effectiveness, I propose a novel, theory-informed approach for generating clinical performance feedback: automated feedback message tailoring. I aim to develop and evaluate this approach

by exploring the potential impact that a knowledge-based, automated feedback message tailoring system could have on clinical performance in low-resource settings.

1.3 SIGNIFICANCE OF THIS RESEARCH

Understanding when AF is effective is increasingly important because of its broad use as a behavior change intervention¹⁰, and the increasing availability of biomedical data that can be analyzed to provide insights into clinical practice.⁸ The significance of this research also arises from the unprecedented growth in the adoption of electronic health information technology (eHealth).^{5,6} With our growing understanding of how and when electronic clinical data can be used for clinical research and quality improvement,²⁰ the potential impact of automated feedback message tailoring is also increasing. Automated tailoring of clinical performance feedback is a novel approach that may benefit many stakeholders, including healthcare providers, clinical supervisors, and clinical researchers. Healthcare providers could benefit by receiving feedback that is more useful and relevant to individual clinical behavior change. Clinical supervisors could benefit by understanding how to provide more useful feedback through the use of a menu of tailored feedback messages. Clinical researchers could benefit by gaining an ability to observe when and how feedback is tailored, thereby creating the possibility to learn about its effectiveness in clinical trials. These benefits may be especially significant in low-resource settings, where performance feedback is limited and a growing presence of eHealth provides opportunities to glean meaningful performance feedback from electronic medical records and other sources of clinical data.

2.0 E-HEALTH IN LOW-INCOME COUNTRIES

The term eHealth is used in this research to mean “the use of information technology in the delivery of healthcare.”²¹ I use this broad definition of eHealth to accommodate diverse forms of information technology used within healthcare settings, all of which generates clinical data that might be used for monitoring clinical performance and generating feedback. Examples of eHealth include electronic medical record (EMR) systems, electronic patient tracking or clinical registry systems that monitor groups of patients with a specific disease, human resource information systems used in hospitals, patient registration systems, laboratory information systems, pharmaceutical inventory systems, web-based and mobile health information systems, and disease surveillance or monitoring systems for public health agencies. The data collected and used in all of these systems could potentially be used for automated clinical performance feedback.

The last decade has seen unprecedented growth in the adoption of eHealth^{5,6}, and this expansion includes low- and middle-income countries. A 2010 systematic review of eHealth evaluations in developing countries identified 45 studies with a total of 55 evaluations of eHealth systems.⁷ The review identified the following categories of eHealth systems: electronic health records, laboratory information management, pharmacy information, patient registration or scheduling, monitoring and evaluation, clinical decision support, patient reminder or notification, and research/data collection. This study excluded telemedicine systems, another widely used eHealth domain. The authors identified 15 qualitative evaluations, and 40 quantitative evaluations. Although the authors determined that little evidence exists about the effectiveness or impact of eHealth systems in developing countries, they identified a recent trend towards increased numbers of studies and increased numbers of randomized controlled trials, suggesting a growing evidence base for eHealth in developing countries.

OpenMRS, an open-source electronic medical record system (EMR) platform is an example of an EMR that is widely used in the developing world.²²⁻²⁴ In 2011, OpenMRS was deployed and used in more than 40 countries.²⁵ A primary motivating factor for the development and implementation of electronic health information systems is to improve the efficiency of data management at the patient and population levels. Disease epidemics in Sub-Saharan Africa, like the human immunodeficiency virus (HIV) and multi-drug-resistant tuberculosis (MDR-TB), have elicited large-scale public health campaigns that introduce clinical information systems to manage data for monitoring treatment outcomes and forecasting drug demand.²⁶⁻²⁸ Public health data that is aggregated from patient-level records is gaining recognition as an efficient approach to monitoring and evaluation of disease treatment programs. Other factors contributing to the growth of electronic health information systems are their potential to improve the quality of healthcare and support expansion of health services to a national scale in developing countries.^{7,29} The expanding availability of reusable electronic clinical data is an important requirement for automated AF and represents a unique opportunity to improve the quality of care in low-resource settings.

eHealth is increasingly used specifically to support national implementations of antiretroviral therapy (ART) programs. A 2008 survey of electronic medical databases used by ART programs in low-income countries was conducted to assess measures used to improve data quality and follow-up for patients who were lost to ART care.²⁷ The study described 21 ART-specific eHealth systems from 15 countries, demonstrating the breadth of use eHealth in low-income countries to support ART implementation. The survey found significant variability in the practices, system attributes, and resources dedicated to the support and expansion of eHealth systems for ART. Most significantly, the survey demonstrated that data quality is frequently poor, with a median percentage of 10.9% of data missing for six key variables across all sites. The authors of this study call for increased human resources and training to manage data and support eHealth for ART.

2.1 DATA QUALITY

Long-standing challenges to using clinical data for quality improvement and research purposes persist.²⁰ A central challenge in the use of clinical data for performance measurement is poor data quality.^{13,30} The assessment of data quality is critical for the effective use of any routinely collected data.³¹ Data quality is widely understood as a multi-dimensional construct, addressing such features as the accuracy, timeliness and completeness of data,³² but it can be understood from two different perspectives. These are a “fit-for-use” perspective, and an ontological perspective.

Data quality is “fit-for-use” in that its determinants are contextualized, depending on the specific purposes for which the data are being used, from the perspective of the data consumer.^{14,15} For example, for a supervisor or public health official, the quality of clinical data that can be used for the purpose of providing clinic-level performance feedback may be significantly different from the quality of the same clinical data from the perspective of an individual healthcare provider, used for the purpose of providing individualized performance feedback. To determine when data quality is “fit-for-us.”, a data quality assessment process should monitor data quality features based on the data consumer’s expectations of the most useful data, to the extent allowed by available resources.³¹ Data quality features include intrinsic measures of data, such as consistency and reliability, and conceptual measures such as timeliness, which are relevant for a specific context. Being based on varying trade-offs and provider expectations, data quality assessment is a subjective and highly contextualized process, but one which is critical for the successful use of eHealth data.

Data quality can also be conceived of from an ontological perspective, meaning that the quality of data is determined not just by the degree to which data are free of error and complete within a database, but by the degree to which an information system represents a real-world system.³² From this perspective, completeness of data reflects the degree to which an information system was designed to portray a complete picture of a real world system. For example, consider a patient as a real-world system, and an EMR as being designed to portray all of the care that a patient has received. EMRs are recognized to be inherently limited in representing the patient as a real-world system, given that patients seek care in multiple

facilities whose record systems are not linked, and even within a single care facility, documentation is often incomplete.¹³ Therefore the use of EMR data to make inferences about care will necessarily portray an incomplete picture of the patient and involve uncertainty. The amount of uncertainty involved in using EMR data will decrease as the representation of the care received increases. Improvement of data quality from this perspective is design-oriented rather than data-oriented. Instead of engaging in data cleaning and routine data quality assessment, from an ontological perspective, improvements to data quality are achieved by changing the design of the information system to improve the representation of the real-world system.

Achieving adequate data quality is primary challenge for the use of eHealth data in low-resource settings. However, I view poor data quality as a barrier that is gradually reduced as our experience in implementing eHealth grows and as our ability to analyze data clinical data to monitor performance improves. While it may not be possible to reliably use eHealth data to generate individualized performance feedback for all providers who use eHealth across a wide range of performance measures, I anticipate that we can develop tools that opportunistically identify data to provide meaningful performance feedback for individuals, and that this work can provide a starting place for the development of systems that ultimately provide highly reliable performance feedback. I view the work that has been accomplished to implement eHealth in Malawi as one of the earliest opportunities to implement individualized performance feedback at a national scale, based on an EMR that is used for the provision of antiretroviral therapy in HIV/AIDS clinics.

3.0 SETTING: MALAWI

Malawi is a landlocked country in Sub-Saharan Africa with a population of close to 17 million people, and a land area equivalent to Pennsylvania. The country has a largely agricultural economy, a highly rural population, and high rates of poverty, with approximately 74% of the population earning less than \$1.25 per day. Malawi has a significant dependency on foreign aid to support economic development, with healthcare in particular being largely supported by donor funding. An estimated 90% of all medication costs are covered by foreign aid.

Like most low-income countries, Malawi has a significant shortage of healthcare providers. The global distribution of healthcare workers (HCWs) is skewed away from low-income countries that hold the greatest proportion of the global disease burden, resulting in a critical human resource shortage in global healthcare.³³ For example, the World Health Organization (WHO) estimates that the disease burden in African countries represents 24% of the global disease burden, while the region has approximately 3% of the world's HCWs.³⁴ Malawi is representative of the broader Sub-Saharan African region in experiencing concurrent disease epidemics. With a ratio of approximately one physician for every 50,000 inhabitants, Malawi and neighboring Tanzania have the lowest doctor-to-patient ratio in the world.³³

3.1 TREATMENT OF HIV/AIDS

The adult HIV/AIDS prevalence in Malawi is approximately 10.8%, with an estimated 1,100,000 people living with HIV in 2012.³⁵ HIV prevalence has gradually declined in Malawi since peaking nationally at 26% in 1998.³⁶ Since the Ministry of Health ART Program began implementing a free, national Antiretroviral Therapy program (ART) in public hospitals in

2004, Malawi has successfully scaled-up treatment to 71% of the population in need of ART. By the end of March, 2014, a total of 486,795 out of an estimated 680,000 patients in need were alive and on treatment, receiving ART from one of 694 sites in the country.³⁷

Malawi’s approach to implementing a national ART program is highly standardized, simplified, and public-health focused. This approach was adapted from the DOTS (directly observed treatment, short course) tuberculosis control framework, which requires limited diagnostic information, limited treatment options for patients, and simplified reporting and drug procurement practices. The implementation approach chosen by Ministry of Health required all public and private ART sites to commit to providing a single first-line fixed dose regimen, to follow national treatment guidelines, and comply with a nationally standardized monitoring and evaluation process.³⁸ This simplified public health approach to ART implementation has been influential among other low-income countries implementing ART.³⁹ Malawi’s ART program centrally supervises and coordinates drug procurement, formulary, treatment guidelines, provider training, monitoring tools, free provision of ART services, and clinical mentoring among other activities, all of which follow a standardized and simplified approach.^{40–42}

National supervision efforts in Malawi are a time-intensive effort conducted on a quarterly schedule that involves a full review of data quality and treatment practices for each site. For the quarter ending in March, 2014, a total of 72 supervisors spent a combined 1,931 working hours visiting 689 public and private healthcare facilities across the country.³⁷

3.2 NATIONAL ART EMR

To support monitoring and evaluation for Malawi’s ART program and improve management of clinical records, the Ministry of Health partnered with Baobab Health Trust, a Malawian NGO, to develop and implement an EMR that could be used as an electronic patient registry, to generate quarterly cohort reports for national supervision. The EMR is a point-of-care, touchscreen-based system that was first implemented at Lighthouse Trust in Lilongwe in 2002 and has since been expanded to additional care programs in more than 50 healthcare

facilities in Malawi.²⁶

The National ART EMR is designed to be used by healthcare providers and staff who collect data using touchscreen computer workstations at the point of care. Using the National ART EMR, health workers enter clinical signs, symptoms, diagnoses and prescriptions in structured formats that are collected in accordance with the standard ART workflow. The data collection process guides health workers through clinical protocols in accordance with Malawi's national ART guidelines. The National ART EMR provides a minimal past medical history, alerts and reminders that encourage adherence to guideline recommendations, clinical calculations such as body mass index, and point-of-care ordering and prescribing. A typical EMR site includes three point-of-care workstations connected to a small server over a local area network. The most common type of the National ART EMR site is an ART clinic within a district hospital, staffed part-time with one or two clinical officers, two nurses, and one registration clerk.²⁶

National ART EMR stakeholders in Malawi have differentiated, valid purposes for the use of EMR data as data consumers. For the purpose of national, regional, or organizational program monitoring and evaluation in clinical settings, the data consumers are administrative supervisors, such as Ministry of Health officials, regional health system administrators, and hospital administrators who require aggregate data in reports showing critical changes in performance to understand the impact of inputs such as pharmaceuticals, training, and supervision activities on clinical processes and outcomes. Such reports are compiled on a quarterly schedule, and while important, are less relevant to individual healthcare providers who are rotating through a clinic, or for supervisors looking to provide individualized performance feedback. The degree to which this data can be used to generate individualized performance reports, to my knowledge, has not been studied prior to this research.

4.0 IMPLEMENTATION SCIENCE

Implementation science is the study of processes of integrating evidence-based practice within a setting.⁴³ Different terms that have been used to refer to implementation science include the following: knowledge translation, research utilization, knowledge transfer, and “dissemination and implementation.”⁴⁴ The field of implementation science can be understood as work that addresses limitations for the processes of knowledge creation, such as publication of original research findings, systematic review of publications, and the publication of that evidence, which are not sufficient to influence clinical decision making on their own.⁴⁵ Implementation science is regarded as broader than clinical translational research in that it addresses the implementation of knowledge into various levels, including biomedical, clinical, and policy knowledge. For example, implementation science is considered to include knowledge about patient experience and preferences, which could be used to form an evidence base to be implemented at a policy level, and this work may be valuable for healthcare systems, but may not directly involve clinical processes.⁴⁵

In the past decade the emergence of the field of implementation science has led to the formation of many models defining the constructs that determine the success or failure of the implementation of knowledge in healthcare. Three notable frameworks that are relevant to the implementation of Malawi’s ART program are the Knowledge to Action Cycle, the Consolidated Framework for Implementation Research (CFIR) and the Implementation Science framework used by the US President’s Emergency Plan For AIDS Relief (PEPFAR).

4.1 THE KNOWLEDGE TO ACTION CYCLE

A widely-used conceptual model for implementation science is the knowledge-to-action (KTA) cycle (Figure 2).^{45,46} The KTA cycle is composed of a knowledge creation process at its center, with seven phases surrounding it, called the action cycle. The knowledge creation process contains *knowledge inquiry* (including the publication of original research findings), *knowledge synthesis* (including systematic reviews and meta-analyses), and *knowledge tools/products* (including clinical practice guidelines and decision aids). The action cycle has seven phases that can be conducted concurrently or individually, each addressing a domain of planned actions that support the implementation of knowledge, based on a distillation of many related theories that concern bringing change into healthcare systems. The KTA cycle is useful as an organizational model for the understanding relationships between various kinds of implementation research, and can be useful as a guide for planning implementation activities.

Considering the KTA cycle in the context of the Malawi Ministry of Health’s implementation of the national ART program is helpful for understanding the work of the Ministry of Health (MoH). MoH is engaged in implementation activities that span the knowledge creation and action cycle processes. MoH’s knowledge creation processes are both *knowledge inquiry* in the form of publication of original research findings about the provision of ART in Malawi, and *knowledge tools/products* as it adapts clinical practice guidelines developed by the World Health Organization for the local context, given available resources and capacity to provide the best care. Another example of *knowledge tools/products* developed specifically by MoH are job aids, such as checklists, charts and diagrams that can be used as quick-references to support the implementation of its national treatment guidelines. MoH is involved in many of the action cycle processes simultaneously to support the provision of ART. These activities include training programs, which fall under *sustain knowledge use*, and its routine supervision efforts that support monitoring and evaluation, which belong in the *monitor knowledge use* and *evaluate outcomes* phases of the action cycle.

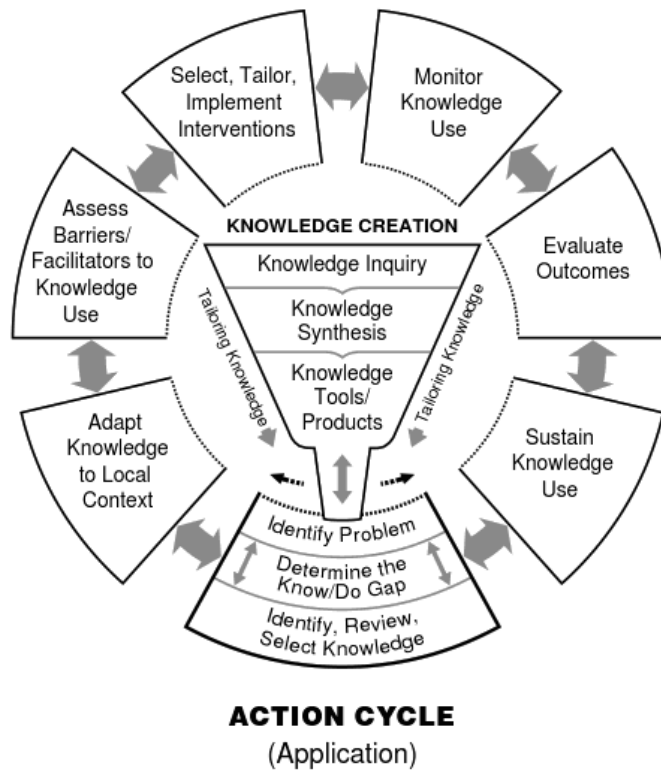


Figure 2: The knowledge to action (KTA) cycle (Graham et al. 2006).

4.2 CONSOLIDATED FRAMEWORK FOR IMPLEMENTATION RESEARCH

The Consolidated Framework for Implementation Research (CFIR) is a meta-theoretical framework for health services research that was derived from 19 published theories in the field of implementation science, with a goal of organizing what is known about implementation within a unified theory.⁴⁷ CFIR consolidates overlapping theories by defining shared constructs that can represent components from multiple implementation science theories. Beyond organizing implementation science knowledge, the creators of CFIR intend it to guide formative evaluations of implementation interventions. CFIR can also be used to interpret research findings relative to other theoretical frameworks. For example, researchers in Kenya who implemented a multi-faceted performance improvement intervention in eight district hospitals mapped their findings to CFIR to demonstrate and understand the generalizability of their results.⁴⁸ The theoretical constructs contained within CFIR are organized into five domains, including the *intervention*, *outer setting*, *inner setting*, *individuals involved*, and *implementation processes* (Figure 3).

The *intervention* domain of CFIR refers to the characteristics of activities that are being put into routine practice, including the core elements of the intervention that can not be compromised, and the adaptable components that can be changed to accommodate the needs of a specific environment. The intervention domain contains the following attributes: *intervention source*, *evidence strength and quality*, *relative advantage*, *adaptability*, *trialability*, *complexity*, *design quality and packaging*, and *cost*. Each of these attributes is defined in terms of the original theories they arise from to enable researchers to understand their applicability to an intervention. Examples of interventions from Malawi’s ART program include the prescribing of new drug regimens, as well as the dissemination of clinical practice guidelines describing the circumstances under which a drug regimen should be prescribed. The intervention includes the ART program’s initial scale-up and the successive changes that occur as the Ministry of Health adapts the treatment program to accommodate new drug regimens, new medical evidence, training needs, and other dynamic influences.

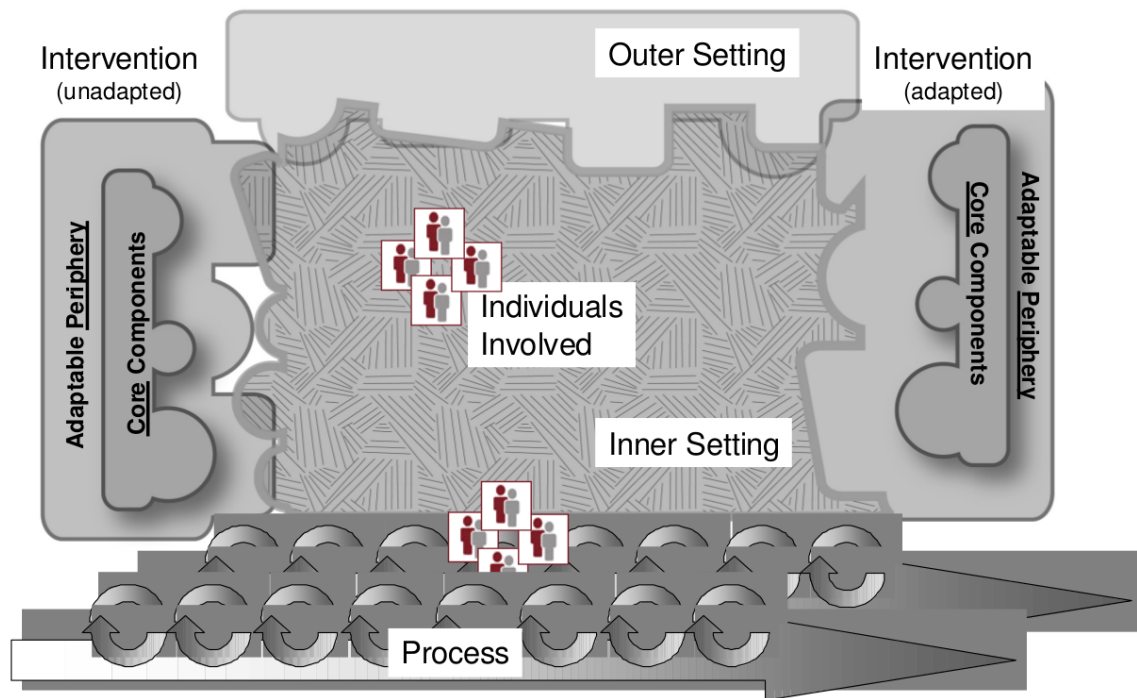


Figure 3: Consolidated Framework for Implementation Research (Damschroder et al. 2009).

The *outer setting* domain of CFIR contains the external aspects of the organization that influence the implementation. This domain includes patient needs and resources, cosmopolitanism (the degree to which the organization is networked with other organizations), peer pressure, and external policies and incentives. In the case of Malawi’s ART program, the outer setting appears most relevant at the level of the ART clinic, each of which differs across the constructs of the outer setting domain. For example, some ART clinics are in close proximity to an urban, clinical center of excellence, while others are in remote locations where access to clinical expertise is more limited. This distance can influence the patient needs and resources, cosmopolitanism, and peer pressure that health workers and clinic administrators experience at the organizational level.

The *inner setting* domain of CFIR can overlap with the *outer setting* domain, but the *inner settings* is oriented toward internal aspects of the organization across multiple levels. The *inner setting* includes *structural characteristics*, *networks and communications*, *culture*, *implementation climate*, and *readiness for implementation*. Each of these constructs uniquely impacts the degree to which an intervention can be successfully implemented, and can vary within the organization. In particular, *implementation climate* and *readiness for implementation* are important constructs in the *inner setting* domain. *Implementation climate* is defined as “the shared receptivity of involved individuals to an intervention and the extent to which use of that intervention will be rewarded, supported, and expected within their organization.” *Implementation climate* contains six sub-constructs, including *tension for change*, *compatibility*, *relative priority*, *organizational incentives and rewards*, *goals and feedback*, and *learning climate*. *Readiness for implementation* contains three sub-constructs: *leadership engagement*, *available resources*, and *access to information and knowledge*. In Malawi’s ART program, the implementation is driven externally to the individual ART clinics, but must accommodate the variability within the *inner setting* of each ART clinic. This construct can guide evaluation of the likely success of an intervention within an ART clinic by providing a comprehensive list of dimensions within the *inner setting* that can facilitate or prevent a successful implementation.

The *individuals involved* domain describes attributes of the people expected to change their behavior as a result of the intervention. This domain includes *knowledge and beliefs*

about the intervention, self-efficacy, individual stage of change, individual identification with the organization, and other personal attributes. Individuals involved draws attention to individuals' differences in terms of personality, skills, knowledge and group interaction within an organization that can represent barriers or facilitators of implementation. For example, the construct *individual identification within an organization* includes the work attitudes and emotional exhaustion of an individual or group. Malawi's ART program requires health workers to change behavior, but some individuals in some clinics may be experiencing increased feelings of emotional exhaustion as a result of increased patient burden. Recognizing individual differences within the ART clinic staff enables the treatment program to be better adapted to accommodate variability across ART clinics.

Finally, the *implementation process* domain contains four activities that are foundational processes of implementation. These are *planning, engaging, executing, and reflecting and evaluating*. These activities are iterative and may be performed in a non-sequential manner. *Planning* refers to creation of a course of action for implementation that is guided by stakeholders' needs and involves defining procedures for tracking progress, training requirements, and piloting or test cases of the intervention. *Engaging* describes the selection of individuals who will support the implementation within levels of the organization and external to the organization. *Executing* refers to the quality of the implementation in terms of fidelity to the implementation plan, timeliness, and degree of engagement with individuals participating in the implementation. *Reflecting and evaluating* are the process of reviewing feedback or metrics about the progress and quality of the implementation, including time for personal reflection to promote shared learning about the implementation. An example from Malawi's ART program that falls within the *executing* construct is its monitoring and evaluation activities designed to provide comprehensive, quarterly progress reports at the national level.

By creating a comprehensive theoretical map, CFIR attempts to unify implementation theory. As such, CFIR provides an opportunity for researchers to identify relevant factors that have been studied as barriers or facilitators of successful implementation across a wide range of implementation settings. In contrast to CFIR, the PEPFAR implementation science framework is being developed as a specialized framework for implementation knowledge about

HIV/AIDS treatment in low-income countries. Relative to CFIR, the PEPFAR framework contains a smaller set of constructs that are couched in the assumptions of a public health approach to ART implementation.

4.3 PEPFAR IMPLEMENTATION SCIENCE FRAMEWORK

The President’s Emergency Plan For AIDS Relief (PEPFAR) is an international HIV/AIDS treatment effort that has supported the implementation of ART in more than 32 countries. PEPFAR adopted an implementation science framework to improve its programs’ development and effectiveness. The primary goals of adopting the framework were to increase the rigor used in evaluating the impact of its programs, and to improve the ability of its partners and other ART implementers to share knowledge. PEPFAR’s implementation science framework is facilitating the transition from an emergency response approach to the HIV/AIDS crisis towards a longer-term strategic approach that many low-income countries are now facing.³⁹ The framework contains three components: *monitoring and evaluation*, *operational research*, and *impact evaluation*.⁴⁹

Monitoring and evaluation is a central activity for public health treatment programs such as Malawi’s ART program. Monitoring refers to the routine tracking of performance at multiple levels of the program that informs progress at regular intervals. Evaluation in this context asks what has been accomplished and measures the benefit of the program to the intended recipients of the programs’ services. Malawi’s ART program has a monitoring and evaluation process that requires quarterly reporting of key program metrics, tracking of drug stocks and patient treatment outcomes. This quarterly reporting process requires a significant effort on the part of staff at all levels of the program in order to collect and maintain the data to determine the ART program’s status and quality.

Operational research, also called operations research, is a “learning while doing” process of using scientifically rigorous research methods to identify and improve implementation once the implementation process is underway.³⁸ Operational research questions identify implementation problems or inefficiencies and their solutions as they arise. Examples of an

operational research question are “What is the best prevention package to reduce mother-to-child transmission of HIV?” and “What is the best method to reduce early mortality for patients starting ART in Sub-Saharan African countries?” Operational research has contributed to the success of Malawi’s ART program in answering research questions regarding treatment outcomes for different patient groups, the adequacy of the program’s data quality, and to understand the causes of patients who were lost to treatment follow-up.⁵⁰

Impact evaluation is designed to make causal attributions about the effect of the program and therefore requires the most rigorous scientific methods. Impact evaluation also attempts to answer questions that address control groups or to estimate the counterfactual, which is the hypothetical state of a population if the intervention had not taken place. This type of evaluation includes randomization or pseudo-randomized approaches such as the “stepped wedge” controlled trial that measures impact in phases as implementation occurs at a small number of sites within each phase. A 2008 impact evaluation of Malawi’s ART program found that patients on treatment between 2004-06 maintained high survival rates one year after starting ART.⁵¹

The KTA, CFIR, and PEPFAR frameworks emphasize similar processes in the implementation of knowledge to improve clinical practice. A primary difference between the PEPFAR framework and the other two are that PEPFAR’s framework is designed specifically to guide ART implementation efforts in low-income countries, whereas CFIR and KTA model implementation science constructs more broadly. KTA and the PEPFAR framework are similar in their emphases on process and action, whereas CFIR appears to focus to a larger degree on context and stakeholder differences. All three models have the potential to inform and organize knowledge about the current activities being carried out by Malawi’s Ministry of Health and for future plans to sustain and adapt the ART program within Malawi.

5.0 IMPROVING HEALTHCARE IN LOW-RESOURCE SETTINGS

Implementation science frameworks are increasingly used to impact healthcare delivery in low-resource settings. Much of the work addressed in the implementation science literature could be considered to overlap with the managerial approaches used in health services and as a reframing of existing practices that have already been used for decades by managers to improve healthcare services, such as training and supervision approaches. As such, these practices are not to be considered relatively novel, but rather as established processes that may be coordinated to implement best-practice knowledge for a specific domain. In low-resource settings processes include task-shifting, supportive supervision, and clinical practice guideline development and implementation.

5.1 TASK-SHIFTING

Task-shifting is a longstanding practice in low-resource settings that aims to enable health workers to safely provide care without the availability of physicians and other specialized cadres of healthcare workers.⁵² The goal of task-shifting is to expand healthcare services while maintaining the quality of care, despite having a shortage of health workers. Task-shifting includes range of practices and differing degrees of implementation in low-resource settings. In a typical task-shifted scenario, no physician is routinely available in the clinic. Physician tasks such as initial clinical evaluation and prescribing are performed by non-physician clinicians. Non-physician clinicians are referred to alternately in low-income countries as clinical officer (the term used in Malawi's health system), health officer, nurse clinician, medical assistant, physician assistant or nurse officer. Typically, non-physician training programs re-

quire completion of secondary school as an admission criteria and are three years in length, with additional training provided for sub-specialties. The scope of practice for non-physician clinicians commonly includes medicine, minor surgery, obstetrics (in some countries including cesarean section), orthopedics, and ophthalmology.⁵³ Nurses are assigned tasks traditionally performed by non-physician clinicians, such as prescription refills and consultation for stable patients. A new class of health assistant called a peer educator is created to perform the nursing tasks that require the least amount of professional training. Clinical officers refer patients who are too complex to be managed in the task-shifted setting to a specialist at a tertiary care center.⁵⁴

Task-shifting has the potential to improve access to care and cost-effectiveness of care by making optimal use of the existing skill mix in developing countries. However, significant challenges in the task-shifted clinical environment are maintaining the quality and safety of care, and sustaining HCW motivation and performance.⁵² Evidence supporting task-shifting interventions largely originates from higher-income countries,⁵⁵ but a recent systematic review of task-shifting for HIV treatment in Sub-Saharan Africa concludes that task-shifting can be successful as a rapid means of scaling-up ART.⁵⁶ Task-shifting interventions must have three conditions be satisfied to succeed: appropriate training, effective referral systems, and supportive supervision.⁵⁷

5.2 CLINICAL SUPERVISION AND MENTORING

Supervision in low-resource settings is widely regarded as a critical component of health worker performance improvement interventions, although evidence supporting its effectiveness is insubstantial.⁵⁸ Supervision is not conducted in a uniform way across low-resource settings, but it commonly involves a district-level or regional health officer visiting a primary care clinic to perform problem-solving, review clinic records, and observe care. A 2011 systematic review evaluating the use of managerial supervision to improve health care in low-income countries found only nine studies meeting inclusion criteria, some of which showed a small benefit to the quality of care.⁵⁹ The review concluded that, because the quality of the

studies was determined to be uniformly low, the effect of supervision on the quality of care in low-income countries is currently unknown.

Supportive supervision may have potential to improve HCW performance.⁶⁰ Supportive supervision is a management approach that moves the focus of supervision away from inspection and control, towards enabling HCWs to collectively monitor and improve their performance. The supportive supervisor emphasizes teamwork and process improvement over individual evaluation, engaging HCWs in problem-solving, self-supervision, and shared decision-making.⁶¹ WHO recommends that supportive supervision and clinical mentoring should be central components of any task-shifting intervention, and has developed supportive supervision and clinical mentoring recommendations specifically for ART scale-up in low-resource settings.^{57,62} There is some evidence that involving local staff in identifying and implementing solutions to problems, a defining characteristic of supportive supervision, is a critical success factor for HCW performance improvement.⁶³ A review of 48 published studies of human resource interventions to improve health worker performance in low and middle income countries found that involvement of local staff in identification and implementation of solutions to performance problems was critical to the success of human resource interventions. This and other aspects of supportive supervision overlap with clinical mentoring, which shares the aim of facilitating HCW performance monitoring and improvement.

Clinical mentoring is defined as “the process whereby an experienced, highly regarded, empathetic person (the mentor), guides another individual (the mentee) in the development and re-examination of their own ideas, learning and personal and professional development.” Clinical mentors share some functions with supportive supervisors, but clinical mentors are usually practicing clinicians who can dedicate more time to clinical teaching and case review than a regional or district supervisor.⁶² Malawi’s ART program is implementing clinical mentoring to support ART scale-up using a WHO recommended training program developed by clinical teaching experts.^{64–67} Supportive supervision and clinical mentoring require HCWs to monitor their performance relative to clear standards, typically disseminated in simplified clinical practice guidelines (SCPGs).⁶²

5.3 CLINICAL PRACTICE GUIDELINES

The Institute of Medicine defines clinical practice guidelines (CPGs) as “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.”⁶⁸ Margolis describes a CPG as a “learning map that identifies the core set of clinical, problem-oriented decisions of a discipline that are linked to relevant scientific knowledge and clinical skills.”⁶⁹ This description emphasizes the use of a CPG for navigation of the medical problem space which, as a generalized representation of the best medical knowledge, is meant to be adapted for use in specific clinical settings. CPGs are commonly portrayed as a kind of reference standard, derived from the best medical evidence, that should be augmented through the further development of clinical protocols that support the adaptation of the guideline recommendation to the constraints and opportunities afforded by a specific clinical setting.

5.3.1 Guideline development

CPGs are commonly developed using a combination of available evidence, based on systematic reviews and meta-analyses, and expert consensus. The past two decades have seen a surge in the development of CPGs that is likely due to their potential to improve the quality of care while reducing costs.⁷⁰ A broader reaction to the proliferation of CPGs has questioned their efficacy, calling attention to CPGs’ inability to address non-uniform clinical problems. The potential harms of CPGs include circumstances that fall under conflicting CPG recommendations and conflicts of interest among CPG developers.^{71,72} A resulting scrutiny of CPGs has in part motivated the development of methods and instruments to improve CPG development, knowledge representation, quality, implementability, and evaluation.^{73–76}

In low-resource settings, simplified clinical practice guidelines (SCPGs) differ from CPGs that are developed for high-income countries in their development process, intended use, and diagnostic processes. International public health organizations such as WHO develop SCPGs as reference guidelines that individual countries can adapt and endorse for use within national disease treatment programs such as HIV/AIDS, TB, and Malaria.^{77,78} Efforts to improve

rigor in both the development and national adaptation processes of SCPGs are increasing as a result of a growing recognition that both development and adaptation processes have been largely unsystematic.⁷⁹ In a 2007 qualitative study, Oxman et al. interviewed department directors at WHO, finding that WHO authors of clinical recommendations rarely made use of systematic reviews and instead relied heavily on expert opinion.⁸⁰ A 2011 re-assessment of the use of evidence in WHO guideline development found that a culture change had taken place within WHO, with all recent guidelines citing systematic reviews, or commissioning their development where none existed.⁸¹ Nevertheless, weaknesses in the guideline adaptation processes of individual countries persist. An ethical analysis and qualitative study of national HIV treatment guidelines in Tanzania and Ethiopia found that WHO reference guidelines, which were based on expert opinion, were adopted without adaptation for the implementing country.⁸² A 2011 analysis of adaptation of WHO guidelines by low and middle-income countries from the eastern Mediterranean found that 19 out of 20 national guidelines reviewed contained important inaccuracies or methodological weaknesses in their adaptation. The primary weaknesses identified in the adaptation processes were the exclusion of intended guideline users in the adaptation process, lack of consideration of the implications of guideline implementation, and deficient methods for selection and analysis of the WHO reference guidelines.⁸³

Whereas CPGs for high-income countries are developed for use by specialist clinicians with reliable diagnostic tools, SCPGs are designed to be used task-shifted HCWs who may not have access to diagnostic tools like complete blood counts or viral load tests.⁷⁸ SCPGs are developed for both disease specific treatment and non-specific treatment that may guide clinicians towards identifying the most likely cause of illness within a general patient population. SCPGs target only the highest causes of patient disability and mortality to reduce complexity in clinical decision-making while optimizing the use of scarce resources. This public health approach commonly leads to a higher sensitivity in diagnostic processes and a lower specificity, and requires clear indicators for patient referral when the SCPG does not address the clinical problem at hand.^{62,84}

The development and dissemination of SCPGs are central to the implementation of treatment programs in low-resource settings. The knowledge represented in SCPGs can be used as

a standard for training, performance improvement, and clinical monitoring and evaluation, which in turn support task-shifting and supervision interventions.^{57,62}

An example of an SCPG used in Malawi is the *Malawi Integrated Guidelines for Clinical Management of HIV* (CMHIV).⁸⁵ CMHIV provides a comprehensive, simplified set of clinical recommendations for the management of HIV in outpatient ART and integrated within antenatal care, maternal care, pediatric medicine, and family planning. CMHIV contains guideline recommendations in multiple formats including checklists, flowcharts, and other graphical formats to improve HCWs' ability to use the guideline in daily practice and to implement the recommended practices correctly. CMHIV is based on WHO's 2010 recommendations for ART and the prevention of mother-to-child transmission (PMTCT) of HIV, adapted by the Malawi Ministry of Health to accommodate available resources and its patient population.⁸⁶ The Malawi Ministry of Health conducts a national training and certification program based on CMHIV for new clinicians, and a refresher training for practicing clinicians when a new version of the SCPG is published. CMHIV-based training and supervision are two components within the multi-faceted process of implementing CMHIV in ART clinics in Malawi.

5.3.2 Guideline implementation

Guideline implementation is the process of facilitating the integration of guideline-based knowledge into routine practice in a clinical setting. Early guideline dissemination efforts rested on the assumption that the primary barrier to the uptake of a new evidence-based practice was lack of knowledge, and that once a guideline was made available, providers would incorporate the new knowledge into routine practice. However, recognition that dissemination of guideline documents alone is largely ineffective has increased the importance of CPG implementation relative to the development and dissemination of CPGs.^{87,88} CPG implementation targets multiple influences of knowledge acquisition and behavior change in a clinical environment, typically using a combination of interventions. Multifaceted guideline implementation can include clinical reminders, educational outreach and tailoring for specific health care worker roles, AF, practice facilitators, dissemination of educational ma-

terials, and other approaches. However, no evidence demonstrates a single intervention or combination of interventions to have greater efficacy than any other intervention, therefore the determinants of guideline implementation success or failure are not well understood.

A 2008 meta-review by Francke et al. analyzed 12 systematic reviews of factors influencing guideline implementation.⁸⁹ The meta-review concluded that the evidence base for determinants of success or failure is thin, and that research directly comparing combinations of implementation strategies is needed. The authors found that 10 of the systematic reviews were of low quality, having extensive or major flaws according to the Quality Assessment Checklist for Reviews.⁹⁰ Two of the reviews were of higher quality, having minimal or minor flaws. Most of the reviews excluded non-English language publications and were constrained to a single medical domain. All of the common findings in the meta-review were drawn from reviews with extensive methodological flaws, lessening the significance of all findings.

A common finding of the systematic reviews of guideline implementation indicated that once implementation activities concluded, guideline adherence returned to pre-implementation baseline levels. Francke et al. identified factors influencing guideline implementation and grouped into five categories of attributes of the implementation. The categories were attributes of: the guideline, the implementation strategy, the professionals, the patients, and the environment. Regarding attributes of the guideline, a single common barrier to guideline implementation was the complexity of the guideline. Attributes of strategies that may improve implementation are a) multi-faceted implementation approaches over single-intervention implementation approaches, and b) strategies that involve active participation and are more closely integrated into the clinical workflow, such as point-of-care reminders. With regard to attributes of professionals, awareness of and disagreement with the guidelines, and amount of experience in the workplace were identified as significant influences. Attributes of patients influencing implementation included patient's resistance to guideline recommendations and patients having co-morbidities, which may decrease health professionals' adherence to guidelines. Attributes of the environment that decreased the likelihood of individual guideline adherence were lack of resources in the clinical setting and negative attitudes among peers or superiors. Again, these findings were common only to the lower-quality reviews, making their claims less significant, and leading the authors to conclude

that little can be decisively claimed about factors influencing success or failure of guideline implementation.⁸⁹

The highest-quality systematic review identified by the authors of the Francke et al.'s meta-review is a 2004 review by Grimshaw et al. This review included 235 studies of guideline implementation that contained 309 comparisons of implementation intervention effects. Beyond comparing implementation effects, the authors also compared the overall cost of the guideline development and implementation efforts with the cost savings or benefit of the intervention where cost data was available. Like Francke's meta-review, this systematic review's primary conclusion was that a weak evidence base prevented the authors from identifying preferential guideline implementation intervention strategies or attributes. The low quality of evidence was a result of common methodological and reporting weaknesses including a lack of reported details, contextual factors, and rationale for the intervention, and potential methodological errors such as missing sample size calculations, unit of analysis errors, and, for interrupted time series designs, having intervals that were too frequent or infrequent to adequately account for potential bias. The studies included had a large number of different combinations of multifaceted intervention comparisons, preventing the authors from conducting a meta-regression analysis with adequate statistical power. Regarding the cost of implementations, economic data were reported in less than 30% of studies, and a majority of the studies reporting costs reported only the cost of treatment, leaving only four studies with adequate economic data to permit cost-benefit analysis.

However, the study provided insight into the most commonly evaluated interventions and their relative effectiveness, despite the poor quality of most studies. Measuring the number of comparisons available in the literature, the authors found that the most commonly evaluated single interventions were clinical reminders (38 comparisons), dissemination of educational materials (18 comparisons), AF (12 comparisons), and multi-faceted interventions that included educational outreach (23 comparisons). The authors found that cluster-randomized evaluations of guideline implementation yielded small to moderate improvements in guideline adherence with the following median absolute improvement in adherence for single-intervention comparisons: clinical reminders - 14%, dissemination of educational materials - 8.1%, AF - 7.0%, and multi-faceted interventions including educational outreach - 6.0%. One

important note is that the studies included in the review are from 1998 or earlier, excluding nearly the past decade and a half of implementation research.⁹¹

Like high-income countries, low-income countries do not have a strong evidence base regarding barriers and facilitators of SCPG implementation, with even less evidence available to inform the effectiveness of interventions. A 2005 review by Siddiqi et al. identified common approaches to guideline implementation in low-income countries. The review included 44 publications about guideline implementation research in low-income countries. Of these, no systematic reviews were identified, but eight randomized controlled trials were included. The type of outcomes measured by the studies were either adherence to the guideline or patient outcomes. The authors concluded that the heterogeneity of the studies and methodological problems prevented them from discerning the effectiveness of different approaches to guideline implementation. However, the authors identified the following common approaches to guideline implementation: AF, local consensus development, education and training, educational outreach, educational materials, local opinion leaders, mass media, marketing, reminders, patient mediated interventions, and combined interventions. Other common approaches in low-resource settings are the use of job aids (pictorial or graphical handouts showing clinical algorithms and treatment recommendations), local facilitators, and supervision within multi-faceted implementation strategies.^{92–94}

Although evidence supporting the effectiveness of guideline implementation strategies is thin, studies of higher methodological quality using rigorous methods and providing a study rationale rooted in theoretical models of implementation science have recently appeared. For example, a 2011 study by English et al. evaluated guideline implementation in Kenya for pediatric care using a cluster-randomized trial design informed by models of behavior change from psychological theory.⁴⁸ The study compared two multifaceted implementation approaches using a cluster-randomized trial with intervention and control arms in eight district hospitals. The intervention group included dissemination of evidence-based guidelines, one week of training, job aids, local facilitators to support implementation, external supervision, six-monthly survey with written feedback, and face-to-face group feedback. The group of hospitals referred to as a control group also participated in a guideline implementation intervention of lower intensity that included dissemination of evidence-based guidelines, one

and a half days of training, job aids, and a six-monthly survey with written feedback. The study by English et al. measured changes in 18 performance indicators that included guideline adherence measures for prescribing and clinical assessment, structural changes reflecting availability of resources, and aggregate scores for the quality of care provided. The authors observed significant improvement in 12 of the 18 measures for the intervention group, relative to the performance of the control group, but noted significant variability in performance across the participating hospitals. Evidence for the effectiveness of the intervention is strengthened by the fact that the participating hospitals experienced high staff turnover during the intervention period, such that 18 months after the initial training was provided, an average of only 8% of the staff who received guideline implementation training remained.⁹²

A major component of the English et al. study was a qualitative evaluation of health worker's perspectives of the guideline implementation activities and barriers to guideline implementation.^{95,96} The authors interviewed 29 health workers and used thematic analysis to identify the following ten themes as barriers to guideline implementation:

1. Incomplete training coverage resulting in inadequate knowledge and skills
2. Inadequacy in standard setting and leadership
3. Lack of recognition and appreciation
4. Poor communication and teamwork
5. Organizational constraints and limited resources
6. Counterproductive health worker norms
7. Absence of perceived benefits linked to adoption of new practices
8. Difficulties accepting change
9. Lack of motivation
10. Conflicting attitudes and beliefs

The identification of these themes called the authors' attention to differences in barriers to guideline implementation in high-income countries. The aspects of barriers in low-income countries identified that are not typically found in high-income countries were variability in the acceptance of guidelines across different health worker roles, a lack of demand for evidence behind the new guideline recommendations, a clear impact of resource constraints

on the ability of the health workers to adhere to the guideline, and a desire for payment related to the implementation that fostered poor expectations when none was given. These themes and differences were echoed by another 2009 study investigating reasons for health worker non-adherence to pediatric disease management guidelines in Tanzania, which furthermore identified disagreement with the guideline as a major barrier to implementation.⁹⁷ The findings of both of these qualitative studies highlight the complexity of the interaction between environment, social norms, workplace culture and individual personalities, and their influence on learning and behavior change in the clinical setting. English et al. characterize the understanding of best practices for implementation in African settings as being at the “blank sheet” stage, further emphasizing the need for qualitative research methods.

English et al.’s qualitative study and cluster-randomized trial represent a cutting-edge approach to implementation research in low-resource settings. The authors followed up on this work by conducting a cost-effectiveness study of the implementation.⁹⁸ They found that intervention resulted in a 25% increase in the estimated quality of care in intervention hospitals, at a cost of approximately \$50 per child admission, compared to a cost of approximately \$31 per child admission in control hospitals. Their analysis, which used incremental cost-effectiveness ratios to assess cost per percentage point improvement in the quality of care, found the intervention to be cost-effective relative to other interventions to improve child health in low-income countries.

The authors recognize that a multi-faceted implementation is a complex task that occurs at multiple levels and is shaped by stakeholders at multiple levels of the health system. As such, the understanding of the success or failure of the implementation needs to be negotiated by all stakeholders, and can not be limited to the “mean effect size” observed in the clinical trial. Furthermore the authors use a conceptual framework that relates the study design to industrial/organizational psychology theory and other theoretical constructs that permit their results to be more broadly interpreted and generalized. The authors mapped their findings to the CFIR framework to further demonstrate the relevance of their findings to shared knowledge in the field of implementation science. Finally, English et al.’s approach to guideline implementation is explicitly designed to treat health workers respectfully, taking a participatory and re-educative approach that involves local problem solving and

partnership in changing clinical practice.⁴⁸

5.4 CONCLUSION

Implementation science is an emerging discipline that is generating evidence and knowledge about how to implement best practices in healthcare. The application of implementation science approaches in low-resource settings shows potential to contribute significantly to the optimization of limited healthcare resources and the improvement of patient care. This potential appears to be increased by the growth of eHealth, which creates further possibilities for innovation in the use of electronic clinical data within implementation interventions.

In settings where SCPGs are being implemented and an EMR is used, there may be a significant role for automated interventions, such as automated clinical AF, to facilitate the uptake of knowledge and clinical behavior change by healthcare providers who routinely use an EMR. AF is especially promising as an automated intervention that can be largely software-based, requiring minimal additional resources to support the routine provision of clinical performance feedback in settings where an EMR has been implemented.

6.0 AUDIT AND FEEDBACK

Audit and feedback (AF) is defined as the provision of clinical performance summaries to healthcare providers, teams, and organizations.⁹ The term “AF” is used to describe a range of interventions that vary significantly in clinical context, provider profession, duration, feedback message design, and targeted behavior. Evidence from the most recent Cochrane review, based on 140 clinical trials of AF, shows that AF can significantly improve compliance with desired practice, but that it is unclear which approaches, under which circumstances, will work.¹⁰ Given the relatively limited insights produced by AF trials to date, AF researchers have recently called for a shift towards comparative effectiveness studies, evaluating how and when AF intervention components will work, rather than its overall effectiveness.¹¹ Researchers have also recently argued that the AF research agenda should shift towards the systematic incorporation of psychological theory in the design of trials of AF, noting a lack of theory-informed AF trials and resulting evidence.¹²

In this chapter I first describe AF interventions and discuss a range of examples that highlight the heterogeneity of AF. Next I discuss AF evidence in general, and within low and middle-income countries in specific. Finally I review other AF research of note.

AF interventions include differing components which are used to target diverse clinical behaviors. I use the term “diverse” to indicate both qualitative differences (e.g. heterogeneity) and quantitative differences (e.g. variability). Behavior-related diversity includes categories of routineness, disease-focus, and medical specialization. AF interventions have been used to target routine behaviors individually, such as hand hygiene, test ordering, screening, and referral that are relevant across medical domains. AF interventions also target groups of related behaviors associated with the management of a particular disease, such as the management of diabetes and ischemic heart disease (Figure 4).⁹⁹ Unlike routine be-

haviors and disease-focused behavior groups, AF has been used to target improvement of specialized clinical skills like ultrasonography¹⁰⁰, surgical technique¹⁰¹, and diagnostic mammography.¹⁰² Within a single category of a targeted behavior, intervention components are heterogeneous with regard to approaches to providing feedback, professional roles of targeted providers, and influence on barriers to behavior change.

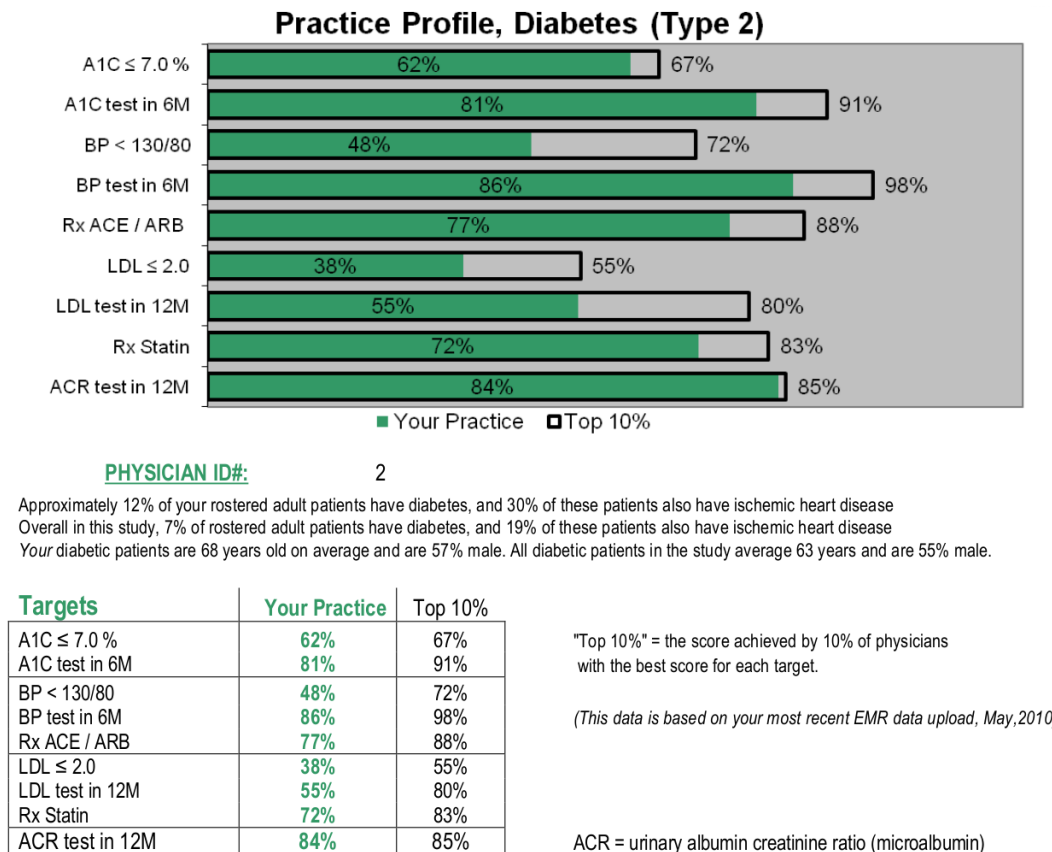


Figure 4: Prototype feedback report for diabetes care used by Ivers et al. 2010.

AF is commonly used to support CPG implementation. A review of guideline implementation strategies found that 24% of guideline implementation studies used AF alone or in combination with other implementation techniques.⁹¹ CPG implementation focuses on increasing individual adherence to best practices derived from the strongest evidence available, provided in the form of guideline recommendations for specific clinical circumstances. When AF is used to support CPG implementation, it is commonly used as part of a multi-

faceted intervention that includes other intervention components. These include educational outreach visits (also called academic detailing), financial incentives, or clinical alerts and reminders.¹⁰³

AF is routinely used for healthcare quality improvement (QI). The QI process is one of the most extensively used approaches to healthcare performance improvement, emphasizing rapid iteration of changes in a clinical setting and monitoring results. QI encourages health workers to ask “What changes can I make that will improve performance?” and “How will I know if a change is resulting in improvement?” without constraining performance measures to be defined according to guideline recommendations. AF supports QI’s cyclical, data-driven monitoring and evaluation process to enable practitioners to determine if changes are effective. AF interventions may be conducted within a QI framework explicitly as part of the Plan-Do-Study-Act cycle (Figure 5), or may be implemented within an implicit quality improvement process that asks the same fundamental questions, but does not specifically use Plan-Do-Study-Act techniques. Audit and feedback conducted within a QI framework is an active process in which health workers themselves typically plan and conduct the measures to be used, data collection and analysis, and feedback delivery.^{103–105}

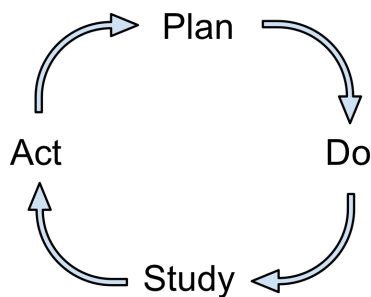


Figure 5: The Plan-Do-Study-Act cycle.

The performance measures used in AF can emphasize *clinical processes* or *clinical outcomes*. *Clinical processes* refer to the intentions and actions of HCWs, such as prescribing a drug, referring a patient, performing an exam, using information tools, using sterile technique, or ordering a test. *Clinical outcomes* refer to clinical end results of processes such

as a patient’s viral load, blood sugar level, or mortality within a patient population.^{105,106} *Process-focused measures* frame clinical performance in terms of the HCWs’ actions and intentions. When performance is based on HCWs’ own actions and intentions, the goal of achieving some level of performance is squarely within the control of the HCW. Guideline implementers typically use process-focused AF to measure performance in terms of HCW adherence to recommendations or protocols, whereas quality improvement practitioners use process measures in conjunction with outcome measures to understand the effect of process changes on clinical outcomes. HCWs receiving process-focused feedback have greater control over the performance outcome, because process measures reflect the HCWs’ intentions and actions. In contrast to process-focused measures, *outcome-focused measures* frame clinical performance in terms of the clinical end results experienced within the patient population. When performance is based on the health outcomes of a patient population, HCWs have less control over the performance outcomes, because many factors can worsen patient outcomes despite the actions and intentions of HCWs.

The data sources used in conducting AF differ in terms of *temporality*, *medium*, *primary use*, and *creator*. *Temporality* refers to retrospective data collection, such as in a medical chart review, or prospective data collection, which occurs during the clinical encounter. The *medium* of the data source can be paper-based records or electronic records. The *primary use* of data analyzed for AF can be as records for patient medical charts, laboratory, pharmacy, treatment registers, public health reporting, or in the case of prospective data collection, the primary use may be for audit itself. Finally the *creator* of the data may be a health worker, a supervisor, or some other administrative staff, each of whom may have varying perceptions and goals within the clinical processes that occur in the workplace. Each of these dimensions may influence the fitness-for-use of the data for AF and thereby impact the effect of feedback on performance.

Feedback features refer to the presentational attributes that convey performance information within a feedback report. Feedback features can vary in terms of aggregation level, confidentiality, social comparison, velocity, correct solution information, frequency, and delivery format. Aggregation level refers to the provision of feedback about the performance of an individual or a group. Confidentiality is the provision of feedback about an individ-

ual performance to only the individual who performed tasks, or to others who may or may not have shared responsibility for the performance. Social comparison, also called normative feedback, refers to the inclusion of performance information about one’s peers compared with group or individual performance information. Benchmarking is a kind of social comparison in which one’s performance is compared with the highest performers within a population ranging from a local to a national level. Velocity refers to the inclusion of data showing performance changes over time. Correct solution information informs the feedback recipient about what can be done to improve performance. Frequency refers to the number of feedback reports that are provided within a specified time period, and can range from bi-weekly to annually. Delivery format refers to both the medium through which the feedback is presented and the means by which the information is conveyed. Delivery formats include verbal, written, computer-based, tabular or graphical display, group or individual presentation, and customizability.^{107,108} Figure 4 shows a prototype feedback report that uses benchmarking, displaying an individual’s performance in comparison with the top 10% of peer performance, presented in both graphical and tabular form. This prototype does not include velocity feedback or correct solution information, but includes both process measures such as “A1C test in 6M” and outcome measures such as “A1C \leq 7.0”.

The nature of the task or behavior that an AF intervention addresses in process-focused feedback may also significantly influence the effect of feedback on performance. Tasks may require team coordination, or may be performed independently by an individual. The frequency with which the task is performed can impact the appropriate reporting frequency. Finally, some tasks require dichotomous measures, indicating whether or not the task was performed correctly, while other tasks require continuous measures, indicating the total number of performances, depending on the nature of the process and the goal.

6.1 EXAMPLES OF AUDIT AND FEEDBACK INTERVENTIONS

To illustrate the range of interventions included under the AF umbrella, I discuss a sub-domain of AF research, which is AF targeting antimicrobial stewardship behaviors. Then I

describe process-focused feedback in a guideline-focused AF intervention in Lao PDR, and outcome-focused feedback in a critical incident AF intervention targeting maternal mortality in Malawi.

6.1.1 Audit and feedback targeting antimicrobial stewardship

Overuse of antibiotics is associated with the complex phenomenon of antibiotic resistance, which has persisted as a high-priority public health concern.^{109,110} Public health organizations promote the use of multi-faceted antimicrobial stewardship programs to improve clinical outcomes, reduce costs, and reduce the spread of antibacterial resistance.^{109,111} Antimicrobial stewardship programs target antibiotic prescribing behaviors such as encouraging reduction of inappropriate antibiotic prescribing (e.g. inappropriately using a broad-spectrum antibiotic when a narrow-spectrum antibiotics is indicated) or unnecessary prescribing (e.g. prescribing an antibiotic when none is clinically indicated). A WHO report on containment of antimicrobial resistance¹⁰⁹ identified the following barriers to behavior change for antibiotic stewardship:

- Lack of knowledge and training
- Lack of access to information
- Lack of diagnostic support
- Fear of bad clinical outcomes
- Perception of patient demands and preferences
- Economic incentives
- Peer pressure and social norms
- Factors associated with the prescriber's working environment
- Lack of appropriate legislation or enforcement of legislation
- Inadequate drug supply infrastructure

To overcome behavior change barriers, programs may use restrictive interventions, such as requiring approval for prescribing of certain classes of antibiotics, or persuasive interventions like educational meetings and AF. Multi-faceted interventions have been shown to be effective for improving antibiotic prescribing in hospital inpatient settings.¹¹² Public

health organizations advocate for the use of AF as a key behavior change intervention to improve antimicrobial stewardship.^{109,113} Nevertheless, evidence about the effectiveness of AF as a behavior change intervention to promote antibiotic stewardship is inconclusive in both ambulatory and inpatient clinical settings. Systematic reviews of interventions to improve antibiotic prescribing behaviors suggest that the success of interventions depends on the specific prescribing behaviors and specific barriers to behavior change in each setting.^{112,114} Used alone, AF appears to have only small effects on prescribing behaviors.¹¹⁴

AF interventions that target antibiotic prescribing use heterogeneous performance measurement approaches for differing provider roles. Measurement approaches may involve techniques such as retrospective chart audit¹¹⁵, daily monitoring and documentation of antibiotic prescription records by a clinical pharmacist¹¹⁶, or electronic prescribing and reporting tools.¹¹⁷ Performance summaries about antibiotic prescribing may include process and outcome measures of performance. Process measures address prescribing behavior and reflect the intent of the provider. For example, a common process measure is the proportion of patients that the provider prescribed guideline-indicated antibiotics for during a reporting period. Outcome measures reflect the patient's disease state or other results of care that are causally associated with the behavior. For example, a feedback report could include outcome measures showing the proportion of patients having different bacterial outcomes (e.g. eradication, persistence, or super-infection) or clinical outcomes (e.g. cured, improving, no change or worsening).¹¹⁶

Provider roles involved in antibiotic prescribing can differ across healthcare profession and specialization. For example, a study targeting antimicrobial stewardship in a teaching hospital in Australia recognized the importance of the roles of nurses, infectious disease specialists, and pharmacists as potential influences on prescribing behavior for junior and senior physicians.¹¹⁷ In non-academic clinical settings, provider roles may be similarly expected to differ as non-physician clinicians frequently may have antibiotic prescribing authority, for example as nurse practitioners, physician assistants, or midwives.

6.1.2 Process-focused AF: Guideline-adherence AF in Lao PDR

An example of process-focused AF used for guideline implementation in a low-resource setting is the use of AF within a multi-faceted guideline implementation intervention. Guidelines are typically implemented by first distributing guideline documents and providing training on the use of a guideline, then conducting AF along with other implementation strategies in a clinical setting. Guideline implementers perform audit by collecting and analyzing clinical examination, treatment, and prescribing data from medical charts and other patient records. Feedback may be delivered to individual clinicians either in written reports or verbally, or at an aggregate level in group feedback review meetings.

For example, guideline implementers used AF in Lao People's Democratic Republic to implement national standard treatment guidelines for malaria, diarrhea, and pneumonia. The audit was conducted using a weighted-score performance indicator that incorporated examination, treatment, and prescription events occurring in individual clinical encounters. Scores were calculated from data collected on paper forms specifically for the purpose of the audit, during each encounter. The guideline implementers compiled aggregate scores for clinicians and discussed the aggregate performance scores in group feedback review meetings.¹¹⁸ In this case the feedback was an aggregate score of individual task performance within a hospital department, derived from the presence or absence of clinician actions for each patient encounter. Examples of clinician actions include weight recorded, patient history recorded, correct dosage and duration prescribed, and whether or not specific actions within an examination performed.

Process-focused audit for guideline implementation typically measures individual task performance because guideline recommendations address individual clinician actions, but performance feedback may be aggregated and presented at the group level. The use of a scoring process benefits the feedback recipients in that it reduces the amount of performance information to be considered down to a single indicator, but it may also serve to obscure the distance between current performance and the goal for individual tasks, which counteracts the fundamental purpose of providing feedback.

6.1.3 Outcome-focused AF: Critical incident AF for maternal mortality in Malawi

An example of outcome-focused AF used for quality improvement in a low-resource setting is *critical incident audit*, which is used to prevent negative clinical outcomes by identifying failures, analyzing the processes leading to failure, and improving clinical processes. HCWs perform critical incident audit by auditing charts or other medical records to identify critical incidents occurring within a specified time period, collecting associated information about each incident, and preparing a presentation to facilitate causal analysis. Health workers present the incidents to peers who then work together to identify clinical process changes that could be made to prevent future incidents from occurring. For example, critical incident audit has been used to prevent uterine rupture in hospitals in Malawi. HCWs performed a monthly chart review to collect data about the occurrences of uterine rupture, the timing of events for a patient leading up to the rupture, and the resulting patient outcomes.¹¹⁹

Critical incident feedback includes the number of critical incidents that occurred within a specified time period, as well as the data about the associated processes that lead to the occurrence of uterine rupture. Critical incident AF is typically used for group tasks like prevention of uterine rupture or prevention of maternal death, where the goal is focused on team efforts to prevent negative outcomes. The group may work to develop and implement a clinical protocol to prevent the negative outcome. The work of reviewing patient charts and presenting the findings is frequently clinician-led and therefore a bottom-up, participatory activity, rather than a top-down, hospital-administration driven activity. The validity of the feedback may be strengthened by the fact that health workers gather the performance information themselves, and the outcome feedback is based on unambiguously negative events, such as uterine rupture or maternal death. Critical incident AF within obstetrics has been frequently claimed to be effective, although very few RCTs have been done in this area.¹²⁰

6.2 EFFECTIVENESS OF AUDIT AND FEEDBACK

Evidence about the effectiveness of AF has been gathered in systematic reviews and meta-analyses of AF clinical trials. Evidence shows that AF interventions have small to moderate positive effects on adherence to desired clinical practice. Taken at face value, evidence suggests that AF interventions are not especially promising as a tool for CPG implementation. However, more detailed analysis of the large body of AF research reveals a spread of intervention effects, ranging from highly effective interventions to negative effects. This variability suggests that AF interventions are not well understood, and that knowing how and when AF works could lead to significant impact on the quality of healthcare.

The most recent Cochrane review of AF, conducted by Ivers et al. included 140 randomized controlled trials of AF in clinical settings.¹⁰ The authors referred to study outcomes as compliance with desired practice. The performance of healthcare providers prior to an AF intervention is called baseline compliance. The authors measured baseline compliance using the median value of performance for the control and AF intervention groups, as a continuous value ranging from zero to 100%.

The authors measured effect size differently for studies with dichotomous outcomes than for studies with continuous outcomes. Studies with dichotomous outcomes included measures such as the proportion of patients who were managed as indicated by a guideline, and the proportion of providers who complied with desired practice. Studies with continuous outcomes included measures of tests ordered or costs incurred. For studies with dichotomous outcomes, the measure of effect was the absolute difference in performance between the intervention and control groups before and after the intervention, called the adjusted risk difference (RD). For studies with continuous outcomes, the authors use a relative measure of effect that was calculated by subtracting the baseline difference in means from the difference in means after the intervention, and dividing that number by the mean performance of the control group before the intervention.

In 49 studies with dichotomous outcomes, the authors found that AF interventions have a median 4.3% absolute improvement effect (inter-quartile range 0.5% to 16%) on healthcare provider compliance with desired clinical practice. In 21 studies with continuous outcomes,

the authors found AF interventions to have a median 1.3% absolute improvement in desired practice (inter-quartile range 1.3% to 28.9%). Although the median effects of AF interventions are generally small, it is important to note that studies in the upper quartile were above 16% absolute improvement, suggesting that AF can have large effects under certain conditions. Notably, studies in the lower quartile had essentially no effect or negative effects, indicating that it is possible for AF to backfire. These findings are consistent with an earlier systematic review of AF¹²¹, and lend credibility to psychological theories that explain the mechanisms by which feedback interventions can be detrimental to performance in some cases, which I will discuss in the next chapter.

To identify features of AF interventions associated with greater effectiveness, Ivers et al. conducted sub-group analyses for a set of potentially explanatory variables across studies with dichotomous outcomes. They conducted sub-group analyses using visual analysis supplemented with meta-regression of a shared effect size, weighted according to the number of healthcare professionals involved. Studies that did not report baseline compliance were excluded from the sub-group analyses, as were studies that were found to have a high risk of bias according to the Cochrane Effective Practice and Organization of Care (EPOC) criteria. The authors use univariate analyses and multivariate analyses where the number of included studies was large enough. The authors analyzed relationships between effect sizes and the several features of AF interventions, including the feedback source, message format, frequency, instructions included, and professional role of the recipient. Ivers et al. found that AF interventions are associated with significantly increased effectiveness when interventions have the following features:

Format includes both verbal and written feedback

Source is a supervisor or colleagues

Frequency is moderate, greater than weekly and up to monthly

Instructions include goal-setting and the creation of an action plan

Direction of change required is to decrease current behavior

Baseline performance is low (at 25%)

Type of professional practice is prescribing

The sub-group analysis that found significant differences across several intervention features that might be used to improve AF interventions. Beyond the Cochrane review of AF intervention, several other recent studies provide insight into the potential mediating factors that can increase the effectiveness of AF.

A 2006 study of AF by Hysong et al. used qualitative methods to evaluate associations between hospital-level performance and use of AF in Veteran’s Administration (VA) Medical Centers in the US.¹²² The authors interviewed clinical staff and administrators at six VA hospitals that were designated as either high or low-performing institutions. After interviewing participants about the use of AF, the authors used grounded theory to create a model of perceived characteristics of AF that are associated with its use in high performing institutions (Figure 6). The model, called actionable feedback, denotes four features of use of AF within high-performing hospitals: *timeliness*, *individualization*, *non-punitiveness*, and *customizability*. *Timeliness* refers to the frequency with which providers receive feedback. The authors considered frequent feedback to be monthly or higher frequency. *Individualization* refers to the provision of individualized performance information, as opposed to provision of aggregate performance reports for a team or clinic, or at the level of the entire facility. Hysong specifically notes that individualization is relevant for clinical guideline implementation feedback where the individual provider is responsible for each task, such as ordering a test or writing a prescription. *Non-punitiveness* concerns the tone of the feedback delivery, or the larger context in which feedback is delivered, such as a supervisor’s use of a supportive tone. Finally, *customizability* refers to an individual’s or a facility’s ability to tailor the feedback to suit their needs.

This qualitative study provides insights into the use of AF at high-performing VA hospitals, which the model of actionable feedback reflects. It should be noted that the model is drawn from provider and administration perspectives of the nature of feedback delivery, rather than observation of the effect of feedback on performance. An important limitation of the model is customizability, a feature which none of the high or low-performing facilities in the study actually had. Customizability was a theme that emerged from the analysis as a feature that high-performing facilities had expressed interest in, therefore it was justified for inclusion in the model. Hysong’s 2006 study is significant in that it recognizes that the

mixed effects of AF from the literature, unlike prior studies of AF which do not appear to consider the potential for feedback to negatively impact performance.

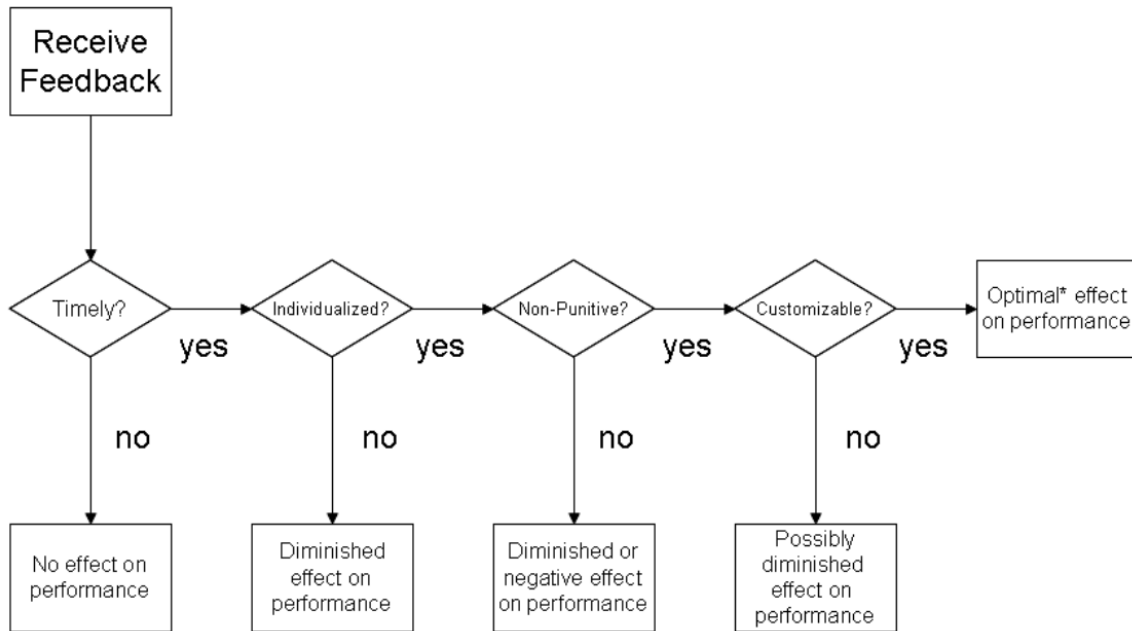


Figure 6: A model of actionable feedback by Hysong, Best and Pugh, 2006.

Following the development of the actionable feedback model, in 2009 Hysong et al. evaluated Feedback Intervention Theory (FIT) as an explanatory theoretical framework for the variability in the effectiveness of AF in healthcare. FIT claims that feedback will be more effective when the attributes and message of the feedback direct the recipient’s attention towards aspects of the task to be performed and away from aspects of the recipient (the self). FIT is described in detail in section 8.3. The authors updated and re-analyzed the 2006 Cochrane review of AF by Jamtvedt et al. using a meta-analysis, to estimate the effect of AF interventions that adhered to or diverged from the practices recommended by FIT. Hysong et al. analyzed the 118 studies included in the Cochrane review, plus three new studies published since the earlier review, using a univariate meta-regression analysis which required included studies have a feedback-only intervention arm compared to a control arm or other intervention groups. This requirement drastically reduced the number of studies

meeting the inclusion criteria to from 118 to 16, plus the identification of three new studies, for a total of 19 included studies. Using an omnibus effect size test, the authors found an effect size of .40, indicating that AF has a modest but significantly positive effect on performance. Their result supported the findings of Jamtvedt et al. both in the effect size and in the fact that five of the included studies had non-significant negative effects, which suggested that the variability could be caused by other unknown moderators of effectiveness.

To evaluate the suitability of FIT as a theoretical framework for AF, the authors tested eight potential moderators of feedback effectiveness that were specifically relevant to FIT. They conducted subgroup analyses using fixed-effects models to evaluate the following factors that FIT would inform: correct solution information, graphical feedback delivery, verbal feedback delivery, written feedback delivery, group vs individual feedback, public feedback delivery, normative information, and feedback frequency. FIT supports the use of correct solution information (“information that helps the feedback recipient see what must change to improve performance”), graphical feedback, written feedback, and group feedback because all of these are likely to direct attention toward the details of the task and away from meta-task processes, or towards the self. FIT posits that feedback delivered verbally, publicly, or feedback containing normative information should decrease the effect of feedback on performance because of the likelihood that these features would direct attention to meta-task processes and self-presentational concerns. Hysong et al., in working with a small sample of 19 studies, found support for three of the eight potential moderators of feedback effectiveness posited by FIT. The authors found that studies providing correct solution information and written feedback reported a significantly larger effect, while studies providing verbal feedback reported significantly smaller effects. They also found one result contradicting FIT, which was that the delivery of graphical feedback significantly reduced the effect of feedback on performance. However, this effect was from a sample of only two studies. The authors concluded, like Ivers et al., that AF can have a modest, positive effect on performance, and that FIT is a viable option as a conceptual framework for the design of AF interventions. The authors also called for future AF research to include more detailed reporting and stricter experimental controls to improve the quality of evidence.¹⁰⁷ Hysong et al.’s meta-analysis stands out as the first known example of the application of psychological theory to the investigation of the

effectiveness of AF and one of the earliest applications of psychological theory to behavior change interventions within the field of implementation science in general.¹²³

Hysong's meta-analysis is representative of a growing recognition in the field of implementation science that clinical behavior change interventions can benefit from psychological theory. Organizational and industrial psychology research has accumulated a wealth of knowledge about the effects of behavior change and performance improvement interventions in workplace settings.^{124,125} To address the theoretical disconnect between implementation science and psychology research, Abraham and Michie developed a taxonomy of behavior change techniques within health care that are linked to psychological theory.¹²⁶ Subsequently, Gardner et al. re-analyzed Jamtvedt's and Hysong's earlier reviews using a meta-regression analysis informed by Abraham and Michie's behavior change technique taxonomy. Using a method to systematically identify behavior change theories to apply to the evaluation of behavior change interventions, Gardner et al. identified control theory as a theory of behavior change that maps most closely with the assumptions underlying AF. Control theory's central premise is that human motivation arises from a desire to reduce a perceived discrepancy between an individual's current state and a goal state.¹²⁷ Control theory is discussed further in section 8. Based on the principles of control theory, Gardner et al. selected the following study inclusion criteria for the meta-analysis of AF studies: a) feedback provided on current performance b) setting of a behaviorally specific performance target (representing a goal state), and c) use of action plans (representing an explicit means to achieve the goal state). Their criteria lead them to find 61 studies which they analyzed using a multivariate meta-regression analysis, to determine the effect of AF on compliance with desired practice. They found that, like Ivers et al. and Hysong, AF had a modest, significant effect on compliance with desired practice (odds ratio ranging from 0.58 to 24.98, median = 1.35, inter-quartile range = 1.02-1.80), thus their control theory-informed analysis did not result in the discovery of a stronger effect within studies that met their inclusion criteria. The authors attributed the lack of a significant result on under-reporting of study details, which prevented them from identifying the use of AF techniques aligned with control theory. This result is discussed further in the context of control theory in section 8. Although the study found no explanatory results about the effectiveness of AF, the authors demonstrated a po-

tentially useful, systematic approach to conducting evidence synthesis that is informed by psychological theory.¹²³

6.3 EFFECTIVENESS OF AUDIT AND FEEDBACK IN LOW AND MIDDLE-INCOME COUNTRIES

Of the 140 studies included in the Cochrane review by Ivers et al., only four were from low- or middle-income countries. To our knowledge, there have been no systematic reviews of the effect of AF in low-resource settings. Two review papers however, both from 2005, examine the effectiveness of health worker performance improvement interventions in low-resource settings, and both evaluate AF among other performance improvement interventions.

Rowe et al. reviewed systematic reviews and other reviews of performance improvement interventions in low resource settings.³ The authors identified 11 literature reviews about performance improvement interventions that mostly address low-and-middle-income countries. Of the 11 reviews, three address AF in combination with supervision. The authors concluded that, based on the systematic reviews they reviewed, they observed a trend indicating that supervision combined with AF is “generally quite effective.” The authors note that multi-faceted interventions are more likely to improve guideline adherence than isolated interventions. However, this conclusion is nearly exclusively based on use of AF for prescribing, rather than other contexts and tasks. While the findings may be relevant for prescribing tasks within ART, they may be less applicable to other types of tasks, such as patient referral or group task performance. The review calls for high-quality research investigating performance improvement in developing countries, noting a lack of rigorous evaluations. Rowe et al. emphasize the importance of understanding when contextual factors may interfere with the generalizability of a study. The authors recognize the importance of connecting interventions to relevant theories in order to build common frameworks that can be used to organize and promote high quality research. They include lists of domains of theories and then specify the interventions that are based on theories within each domain, but list these only at a high-level that does not attribute specific psychological theories to AF. This review

is valuable in its recognition of a need for the use of conceptual frameworks to inform the evaluation of performance improvement interventions.

Another review from 2005, by Siddiqi et al., evaluates the effectiveness of guideline implementation interventions in low-resource settings, specifically addressing AF. In the review, Siddiqi et al. find 15 studies evaluating AF, with one of those studies being an RCT. While almost all of the 15 studies demonstrated an improvement in guideline adherence or patient outcomes, most had design flaws, the authors noted. One important consideration put forth by Siddiqi et al. is the fact that the literature evaluating AF from developing countries is mostly positive, which is not reflected in the conclusions of systematic reviews of RCTs done primarily in high-income countries. The authors speculate that publication bias against negative findings may contribute to this relative lack of ineffective AF evaluations from low-resource settings. Nevertheless, Siddiqi et al. ascribe great potential to AF for low-cost performance improvement in developing countries. Notably, the authors do not question the issue of atheoretical approaches to designing AF interventions, nor do they acknowledge the variability of activities that are attributed to AF.¹²⁸

The majority of publications about AF in low-resource settings over the past decade address either prescribing behavior or critical incident audit for obstetric care, aimed to reduce rates of perinatal mortality and morbidity. Within the critical incident audit literature, few if any studies are RCTs. The fact that a 2005 Cochrane review was initiated to evaluate the effectiveness of critical incident AF in reducing maternal mortality and morbidity in low-income countries is indicative of the growing interest in this intervention technique for developing countries. However, the authors found no suitable trials to include in the review and none have been found in updates of the review through 2011.¹²⁹

Specifically within Malawi, three recent studies from Thyolo District in Malawi's southern region are indicative of a broader trend towards the use of critical incident AF in low-resource settings. Two observational studies measured reduction in negative clinical outcomes using before-and-after studies of obstetric outcomes for critical incident AF. A 2009 study observed a reduction in the incidence of uterine rupture from 19.2 per 1000 births prior to the study, down to 6.1 per 1000 births at the conclusion of the study period.¹¹⁹ A 2011 study at the same district hospital observed a reduction in the incidence of severe maternal complications from

13.5 per 1000 births to 10.5 per 1000 births at the end of the study period. This study noted significant reductions in mortality, hemorrhage, and uterine rupture, and non-significant reduction trends in eclampsia and peripartum infections.¹³⁰ Another 2011 study conducted in Thyolo District used qualitative methods to evaluate health worker perceptions of critical incident AF. The authors concluded that, while a minority of staff had expressed fear about the audit sessions where critical incidents were discussed, most health workers viewed the technique positively and perceived the purpose of AF to be about learning, improving the quality of care, and improving focus and motivation within the clinic.¹³¹ Critical incident AF contrasts with the more standard practice of providing feedback to individual providers about their prescribing or test ordering behavior by actively engaging HCWs in the audit process. The recent publications describing clinical AF in Thyolo district were performed largely in a participatory manner lead by local health workers, rather than as a governmental initiative or other national program to improve the quality of care in Malawi, which may have further contributed to their success. Studies of AF in low-resource settings appear to have generally more positive effects than studies conducted in industrialized nations, but few evaluations of AF in low-resource settings are RCTs, preventing observation and analysis of effect. Observational studies in Malawi in particular have demonstrated significant impact of AF for critical incident audit within the domain of obstetric care.

Evidence addressing the effect of AF on health worker performance shows modest improvement resulting from the use of AF, but with wide variation in effect that includes studies showing performance decreases. This variable evidence belies two issues of note: 1) the practice of AF is heterogeneous, including activities that vary in approach, targeted behavior, professional role, and context and 2) the mechanisms by which performance feedback impacts behavior change are not well understood. The synthesis of evidence about the effectiveness of AF is hindered by its frequent evaluation within multi-faceted performance improvement interventions that do not evaluate the effect of AF alone within a controlled setting. A consensus is forming around the position that future studies of AF should not evaluate its efficacy in general, but rather should investigate the features of feedback as moderators of the effect of feedback on performance within head-to-head comparisons in controlled contexts.^{103,132} Furthermore, increasingly, researchers are turning to psychological

theories of behavior change inform research questions about the underlying mechanisms by which AF impacts behavior.

6.4 CONCLUSION

In summary, evidence shows that AF interventions appear to have a wide range of effects on performance, including large positive effects and null or negative effects for a significant proportion of studies. This large variance in effects on performance may be due in part to the heterogeneity of targeted behaviors, barriers to change, performance measures, AF components, clinical settings, and healthcare provider roles in AF intervention contexts that have been studied. Furthermore, evidence shows that recent clinical trials of AF are not adding to our insight into how and when AF is effective. To improve our understanding of how and when AF works, researchers are increasingly looking towards the application of psychological and behavioral theory to AF approaches.

7.0 USING THEORY TO INFORM IMPLEMENTATION

A theoretical construct is a “concept specially devised to be part of a theory.”¹²⁴ Psychological theories contain theoretical constructs that explain causal mechanisms that are relevant to behavior change processes. Psychological theory offers many credible explanatory causal mechanisms that could be used to understand how to improve AF^{124,133}, but AF research has rarely explicitly used theory to inform intervention design, and no consensus has been established for a theoretical approach to AF research.¹² The use of theory in implementation research has been debated, relative to the merits of pragmatic and empiric approaches.^{134,135} While recognizing the importance of approaches to research that are not explicitly theory-based, I view the explicit use of theory as promising and efficacious for investigating causal relationships between elements of AF interventions, to understand how and when AF interventions are most effective. Three frameworks that concern approaches to the use of theory, and which are relevant to AF interventions, are the Theoretical Domains Framework (TDF), the capability, opportunity, motivation and behavior (COM-B) framework, and the menu of constructs approach.

7.1 THE THEORETICAL DOMAINS FRAMEWORK

Michie and colleagues have argued that psychological and behavioral theory hold significant potential to guide implementation science research towards understanding how and when interventions are most effective at changing behavior.¹²⁴ A central claim made by Michie et al. is that psychological theories that are relevant to behavior change proliferate and have overlapping or shared constructs, making them difficult to identify and apply coherently,

for researchers seeking to use a theoretical basis for a specific intervention. To resolve this issue Michie and colleagues conducted an expert consensus process with psychological theory experts, health services researchers, and health psychologists to develop and validate a framework, called the TDF, for using psychological theory in health-related behavior change interventions.^{124,136}

The TDF is a taxonomy of 13 behavior change theory categories that researchers can use to identify theory that may be relevant to a specific behavior change intervention.¹³⁶ Within each category is a coherent and validated set of theoretical constructs. An example of a TDF domain is “Beliefs about Capabilities” which contains “Self-efficacy”, a construct from social cognitive theory that has been widely studied.¹³⁷ Each theoretical construct within the TDF asserts one or more causal mechanisms that are relevant to behavior change processes. Michie et al. proposed that researchers could use interviewing to identify implementation problems that are associated with a particular domain, and then to further investigate the implications of theories associated with the construct within a theoretical domain. The TDF offers researchers a validated means for selecting a theoretical construct to use to inform behavior change intervention research. To further facilitate the process of systematically identifying theory relevant to behavior change, Michie and colleagues developed the COM-B Framework.

7.2 COM-B FRAMEWORK

The capability, opportunity, motivation and behavior (COM-B) framework for understanding behavior was developed for use within implementation interventions in clinical and health-related contexts (Figure 7).¹³⁸ COM-B models the determinants of behavior, all of which correspond with a specific barriers or facilitators of behavior change: *Capability* refers to determinants such as an individual’s knowledge, skills, and beliefs that create the capacity to conduct a behavior. *Opportunity* contains the environmental influences and other external processes that influence a behavior. An individual’s *motivation* refers to cognitive, emotional, and psychological processes that direct or stimulate behavior. Behavior influences and is

influenced by determinants in the other three categories. Barriers to behavior change are manifested in one or more COM-B category for each individual. COM-B categories have been mapped to the TDF domains to guide researchers in selecting the most relevant TDF domain for a specific behavior change intervention.¹³⁶ For example, the TDF domain of “Beliefs about Capabilities” was mapped to the COM-B “motivation” category. Used together, the TDF and COM-B enable researchers to identify relevant theoretical constructs associated with barriers for a specific behavior that they are aiming to change.

To explore the heterogeneity of barriers to behavior change using COM-B, I discuss examples of antimicrobial stewardship behaviors, informed by a scenario in which a supervisor who is giving verbal feedback to an individual might tailor the feedback in accordance with changes in the environment, or to meet the needs of the individual provider as they receive feedback.

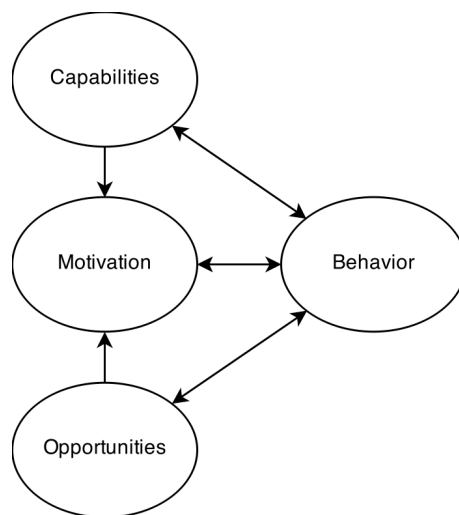


Figure 7: The COM-B framework for understanding behavior (Michie et al. 2011).

7.2.1 Capability barriers

Capability barriers to behavior change refer to the required knowledge and skills that an individual must possess in order to conduct a behavior. Behaviors addressed by AF in-

terventions commonly require individuals to possess multiple, coordinated capabilities. For example, many clinical tasks require both medical decision-making and patient communication skills. Differences in provider training, work experience, knowledge maintenance, and innate abilities can contribute to capability differences.

Supervisors who provide performance feedback may accommodate capability differences by recognizing the set of necessary capabilities, and tailoring feedback messages to address the specific capability they perceive as the most significant barrier to improving performance. For example, reduction of unnecessary antibiotic prescribing requires domain knowledge to recognize the conditions under which prescribing should be delayed, and interpersonal skills to persuade a patient that prescribing antibiotics is not the best action to take. Poor performance in reducing unnecessary antibiotic prescribing could result from lack of either knowledge or skill capability. Consider a supervisor who believes that a low-performing physician has adequate domain knowledge for delaying antibiotic prescribing, but lacks patient communication skills as evidenced by his patient experience survey scores. To address the most likely capability barrier for the low-performing physician, the supervisor might not focus on the negative performance information, but instead reassure the physician about her confidence in his medical knowledge, and recommend training to enable the physician to develop better communication skills. For a high-performing physician, giving feedback about antibiotic prescribing would represent a low-priority task because of the physician's demonstrated competence. As performance improves over time, repeated feedback indicating high performance demonstrates the acquisition of all necessary capabilities, and therefore it loses priority among feedback messages because of its lower informational value and lower potential to change future clinical behavior.

7.2.2 Opportunity barriers

Opportunity barriers are external or environmental constraints on a provider's enactment of a behavior. Behavior in clinical settings has multiple, dynamic opportunity barriers. From an informatics perspective, considering the clinical environment to be a complex socio-technical system¹³⁹, the are following examples of opportunity barriers that are typically not

accommodated by AF:

- *Large problem spaces*: For example, clinical guidelines frequently do not address interaction between multiple medical problems within a patient.
- *Disruptions*: Medical emergencies, infrastructure failure, and disease outbreaks are rarely acknowledged by routine audit.
- *Uncertainty*: Patients presenting with multiple symptoms may lead to diagnostic uncertainty that is not addressed by a guideline.
- *Social influence* from patients and co-workers must be negotiated, is dynamic, and can lead to goal conflict.
- *Automation* can constrain behavior as tools become embedded in the cognitive work of healthcare, yet they may also cause unintended errors.

When a supervisor gives face-to-face feedback to a healthcare provider, the supervisor can interpret performance reports using a wealth of information from the supervisor's own experience of the events that occurred during the reporting period. At best, conventional audit measures accurately represent the environment with regard to a narrow set of information that the individual may not be monitoring. However, even in an ideal situation, there is potential for significant heterogeneity in environmental factors to influence behavior in unpredictable ways. In low-resource settings, opportunity barriers may have more significant influences on behavior. For example, a shortage of antibiotic drugs in a low-resource setting creates a barrier that artificially improves performance for inappropriate prescribing until the drug is restocked.

7.2.3 Motivational barriers

Motivational barriers refer to the internal psychological and cognitive processes that prevent individuals from conducting a behavior. AF interventions address behaviors whose motivational barriers are multi-dimensional and can change from situation to situation, such as beliefs, emotions, intentions, goals, and identity.¹³⁶ Motivation can also affect behavior in response to feedback via influences on how feedback is perceived, acceptance of the message, and desire and intent to respond to feedback messages.¹⁴⁰

For example, emotions have a significant role in feedback interventions as individuals perceive performance feedback through their own emotional and reasoning filters.¹⁴¹ Supervisors who provide performance feedback may aim to emotionally prepare individuals to receive performance feedback. For example, a priming technique called the “feedback sandwich” has been widely used to deliver criticism about negative performance. To make a feedback sandwich for a recipient, a supervisor gives the recipient positive feedback first, then briefly gives the negative feedback, then finishes with another positive message. Research suggests that the feedback sandwich is not an effective technique¹⁴², but the practice demonstrates how supervisors may heuristically frame feedback messages to accommodate recipient emotions.

The examples of barriers to behavior change discussed above that are associated with capability, opportunity, and motivation as determinants of behavior illustrate potentially important implications for AF interventions. Firstly, the behavior change barriers of individual healthcare providers may differ, creating the potential for different barriers to exist among a group of healthcare providers within a feedback intervention. This implication has been recognized by others.¹⁸ Secondly, supervisors have some awareness of the nature of a recipient’s specific barriers to behavior change, and a supervisor may intuitively or heuristically tailor an intervention for the perceived barriers to behavior change that they identify for each individual. Thirdly, tailoring of verbal feedback is potentially a significant hidden mediator, associated with the supervisor, that could influence the effect of AF interventions.

The use of COM-B and the TDF together hold further implications for AF interventions. Associations between perceived barriers to behavior change and COM-B determinants could enable the identification of theoretical constructs, using the TDF, that offer more explanation about the causal mechanisms that make feedback effective for changing behavior. Furthermore, the association of a barrier to behavior change with a theoretical construct identified using the TDF via COM-B might be used to predict the effect of a feedback intervention on performance. The identification of TDF constructs that could be used to inform the design of a feedback intervention have been discussed in a “menu of constructs” (MoC) approach to using theory to design audit and feedback interventions.

7.3 MENU OF CONSTRUCTS

Theoretical constructs from many theories could potentially inform the design of feedback interventions. Because of the heterogeneous nature of AF interventions, which is created by a diversity of approaches, contexts, provider roles, and barriers to behavior change in AF interventions, Brehaut and Eva have argued that no single theory is likely to encompass all of the causal mechanisms that might be relevant for AF interventions.⁹ For this reason, using AF as an example intervention, they proposed that complex interventions to change health-related behaviors use a “menu of constructs” approach, which involves the selection and evaluation of theoretical constructs from many relevant theories to create new representations of a network of causal mechanisms that may mediate the effects of a behavior change intervention.⁹ This approach contrasts with the proposed use of the TDF, which is to identify relevant theory that can be applied as a whole for an intervention. Instead, the authors propose that researchers could select specific constructs that are determined to be most relevant within a particular setting, but might not necessarily need to apply all constructs from a theory to a specific intervention. Using expert consensus, usability testing, or pilot studies, a set of candidate constructs could be identified, and this set could be evaluated rigorously to identify a generalizable menu of constructs that are relevant for a targeted behavior in a particular context.

A menu of constructs approach could be used in conjunction with the TDF and COM-B to identify constructs. However, in discussing the use of the TDF using AF research as an example, Brehaut and Eva raise concerns about the degree to which the TDF excludes cognitive constructs that nevertheless have important implications for behavior change interventions. Characterizing this as a level of abstraction problem, the authors suggest that, because cognitive theories tend to study specific cognitive mechanisms instead of behaviors, cognitive theories were overlooked by the expert consensus process used to formulate the TDF, and most cognitive constructs were collapsed into a relatively small set of high-level constructs, including “memory”, “attention”, and “decision making” that are less useful for understanding many potentially relevant cognitive mechanisms, especially with regard to AF.

7.4 SUMMARY

Researchers have recognized the potential for psychological theory to contribute significant insights to behavior change interventions aimed to improve healthcare in clinical settings. Frameworks have been developed to systematically guide the application of theory to behavior change interventions. Approaches that use the TDF, COM-B, and menu of constructs enable researchers to consider a set of established theoretical constructs that have been widely studied and are likely to be useful in informing the design of a behavior change intervention. Although the TDF has been validated for the systematic identification of psychological theory to inform behavior change interventions, it may exclude theories that are relevant to behavior change processes at lower levels of abstraction (eg at the level of cognitive mechanisms that influence behavior).

In using the COM-B categories to consider the provision of verbal feedback from a clinical supervisor to a recipient, I am calling attention to the existence of individual differences in capabilities, opportunities, and motivation within a group of healthcare providers. The existence of these differences implies that feedback interventions could be improved by tailoring feedback to suit individual barriers to behavior change. This recognition of individual differences is supported by the reasonable expectation that clinical supervisors tailor feedback for individuals, to accommodate individual and situational differences. Tailoring of AF interventions for a group of healthcare providers has been discussed in the AF literature, but to my knowledge, there has been little or no discussion of how to tailor feedback to accommodate individual and situational differences that might change even during the course of the intervention. To explore the possibility of finding theoretical constructs that are relevant to the tailoring of feedback messages but which have not necessarily been included in the TDF, I reviewed the literature on feedback-specific psychological theory.

8.0 FEEDBACK THEORY

Feedback theories offer many causal mechanisms that may be used to inform the design of AF interventions. For the purpose of behavior change, performance feedback is likely to be effective when messages influence a recipient’s specific barriers to behavior change.^{16,17} However, barriers to behavior change differ across individual healthcare providers, stemming from differences in providers’ training, knowledge, work experience, personality and other individual characteristics.¹⁸ Furthermore, barriers to behavior change may be dynamic, as providers’ beliefs, motivations, and perceptions are influenced by ongoing changes in the healthcare organization, the complexity of which is widely recognized.¹⁹

More than a century of psychological research addresses the effect of feedback on performance. Research in the early 1900’s investigating “knowledge of results”, a term analogous to feedback, evaluated the effect of an individuals’ awareness of the outcomes of his or her behavior on task performance.^{143,144} In the following decades, much progress has been made toward understanding how feedback impacts performance, although many unanswered questions about the mechanisms by which feedback impacts performance remain. Competing and complementary theories have attempted to explain the mechanisms of individual and group responses to feedback, and to account for variability and unanticipated effects of feedback interventions on performance. Of the many theories that address the effect of feedback on performance, some of the most influential theories are Feedback-Standard Comparison Theories, the Feedback Process Model (FPM), Feedback Intervention Theory (FIT), and Regulatory Focus Theory (RFT). A smaller body of research investigating group feedback offers insights about feedback provided to groups and team performance. Feedback theories offer a potential wealth of knowledge that has only recently begun to be applied to feedback research and interventions in healthcare. In the following sections I discuss influential

theories of feedback and their relevance for clinical AF.

8.1 FEEDBACK-STANDARD COMPARISON THEORIES

Feedback research within the field of psychology is heterogeneous and addresses a wide variety of tasks, contexts, and types of performance feedback. Despite this variability, most feedback researchers agree that feedback functions as a process of measuring a distance between a goal or standard and a current state of performance, also referred to as a *feedback-standard comparison* (FSC). Two theories that define the *feedback-standard comparison* construct are control theory and goal-setting theory.

The central idea of control theory is that humans are motivated to reduce a perceived discrepancy between their performance and some standard or goal.¹²⁷ According to control theory, behavior is controlled using a negative feedback loop in which an individual works to perform error correction or problem solving to completely reduce the discrepancy. Goal-setting theory on the other hand uses a similar construct but characterizes an individual as determining their goal and then developing performance strategies to approach and achieve a goal state.¹⁴⁵ According to goal-setting theory, individuals are motivated by the desire to achieve the goal, whereas under control theory the source of motivation is discrepancy reduction. Although control theory and goal-setting theory have differences, the fundamental mechanics of feedback-loops within the theories are similar in the following way: a goal has been set, feedback provides information about the distance between current performance and the goal, and motivation arises from the desire to eliminate the distance between the current state and the goal.

FSC theories recognize four different approaches that individuals use to eliminate the gap between the current state and the goal. These are a) to increase effort, b) to abandon the standard, c) to change the standard to be within reach, or d) to reject the feedback message.¹⁰⁸ FSC theories' most important implication for feedback interventions is that, to motivate an individual, feedback should specifically address the gap between the current state and the goal. These theories concern to the source of motivation in response to feedback on

a fundamental level, but do not account for other factors that impact responses to feedback including affective reactions to feedback, the interaction between feedback and learning, and the coordination of multiple goals.¹⁰⁸

Within the AF literature, goal-setting theory and control theory have been used to model the fundamental mechanisms by which feedback impacts performance.^{107,123} For example, if a healthcare provider has a goal of adhering to a guideline recommendation in 100% of relevant clinical encounters, and current performance is below 50% adherence, FSC theories would claim that the perception of current performance being distant from the goal would motivate the provider to eliminate the discrepancy between the goal and current performance. Furthermore, according to FSC theories, the method of eliminating the discrepancy might be achieved in one of four ways, including a) increasing effort or seek new ways to improve performance, b) to decide that the guideline recommendation is not credible, and abandon the goal, c) to decide that the goal of 100% adherence to the goal is not achievable, and set a less demanding goal, such as 80%, or finally d) to decide that the feedback message is inaccurate, and to believe that performance is actually higher than the message indicates. Recognizing that individuals may choose to eliminate the feedback-comparison gap in ways other than increasing effort is an important insight offered by FSC theories for AF interventions.

A important insight provided by FSC theories is that feedback must address a goal that the individual holds, or there will be no perceived discrepancy to motivate an individual. For example, when an AF intervention provides process feedback to improve adherence to a guideline recommendation, if the healthcare provider does not believe that the process will impact the intended outcome, feedback about the process will not result in a perceived discrepancy that would motivate the provider. For example, Malawi's ART guidelines recommend that patients with treatment side effects that do not respond to initial management be referred to a specialist. For patients living in an area that is far from any specialist and who not have the means to pay for transportation and accommodation to reach a specialist, providing feedback to increase rates of referral may not impact the performance of providers due to their belief that referral is not feasible for a large proportion of their patient population.

Another situation in which an HCW may not hold a goal addressed by a feedback in-

intervention is as a recipient of peer or social comparison feedback, which is commonly used in AF interventions as a benchmarking technique. Peer comparison feedback provides an individual with information about the performance of peers relative to his or her own performance. This type of feedback establishes an additional feedback-standard gap in addition to that provided about individual performance. This secondary comparison is only likely to motivate improved performance if the individual has a goal of performing at or above average peer performance. If a peer-performance goal is not held by a provider, social comparison feedback may have little effect.¹⁴⁶ Benchmarking, a variant of peer comparison feedback, displays only a range of the highest performing peers' performance, rather than showing all peer performance. Benchmarking similarly operates under the assumption that an individual will want to achieve a performance level that is among that of the highest-performing peers.

A third important insight offered by FSC theory is that feedback information can be specifically about the recipients' performance-goal discrepancy, or it can be tangential to the specific performance goal. For example, feedback features such as velocity (relative change in individual performance), and delivery format reflect differences in the portrayal of the gap between performance and the goal. Aggregation level of performance events, peer comparison, and correct solution information indicating how to improve future performance are feedback features that reflect the inclusion of additional information which may leverage other mechanisms to improve the effectiveness of feedback, but do not specifically describe the feedback-performance gap.

Hysong's model of actionable feedback, developed using a grounded theory approach to evaluating AF in VA hospitals, is largely aligned with FSC theories. None of the models' constructs, which are timeliness, individualization, non-punitiveness, and customizability emphasize feedback features that would obscure or distract from the feedback-standard gap, and at least two constructs appear to emphasize the gap. The first of these is individualization, which allows a recipient to understand her individual feedback-standard gap, rather than having this information obscured within an aggregate feedback performance including the performance of co-workers. Second, non-punitiveness excludes the addition of additional information in the form of threats or punishment that do not specifically convey information about the feedback-standard gap.

Gardner et al. used control theory in a meta-analysis that measured the effect of interventions using AF, but did not arrive at a significantly different conclusions than other systematic reviews of AF. The authors noted that a lack of detailed reporting in AF studies prevented them from identifying interventions that were aligned with control theory. This is a valid justification of their results because of the complex nature of AF interventions, in which it is possible for multiple aspects of an intervention that are misaligned with control theory to be unreported in a publication about the study.

The low-level mechanisms of responses to performance feedback that are addressed by FSC theory are contextualized for workplace-related motivations and influences on performance by the Feedback Process Model, discussed in the next section.

8.2 FEEDBACK PROCESS MODEL

The Feedback Process Model (FPM) was introduced as an explanatory model of cognitive mediators of the effect of feedback on an individual's performance within an organization by Ilgen, Fisher and Taylor in 1979.¹⁴⁰ The model was validated as a predictive model of cognitive responses to performance feedback in a modified form by Kinicki et al. in 2004.¹⁴⁷ FPM was originally created to organize findings from an extensive review of the feedback literature (Figure 8). FPM models an individual's response to performance feedback as being moderated first by perceptions of the feedback message or *complex feedback stimulus* and by the *source* of the feedback. Next, the effect of a feedback message can be mediated by the following sequence of cognitive variables: *perceived feedback*, *acceptance of feedback*, *desire to respond to feedback*, and *intended response (goals)*. Each variable in the sequence is capable of attenuating the effect of feedback on performance, contingent on the effect of the preceding variable. Each variable in the sequence is also presumed to be influenced by individual difference characteristics such as personality, ability, and motivation. Finally, the recipient's actual response is recognized to be influenced by *external constraints* that represent any possible external influence on an individual's performance. Ilgen et al.'s review discusses each cognitive construct in terms of the feedback source, message, and recipient.

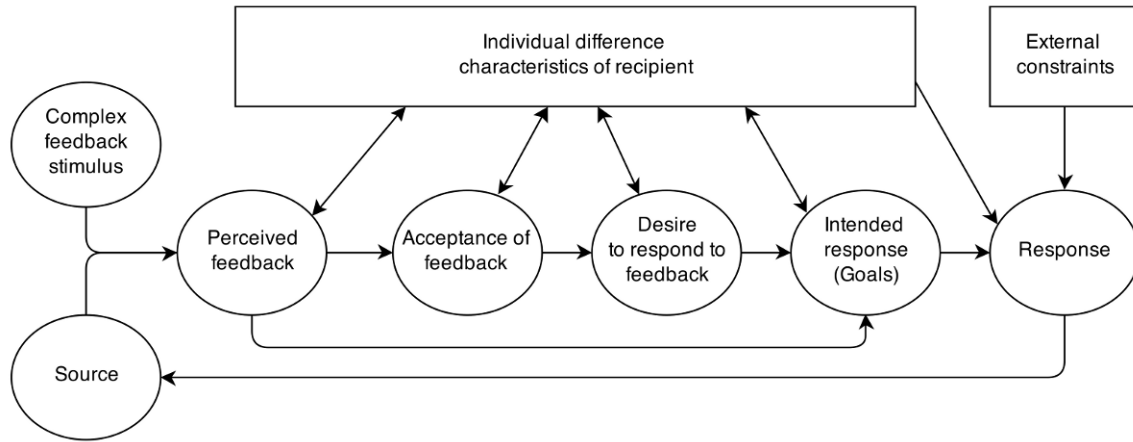


Figure 8: A cognitive model of feedback processing from Ilgen, Taylor and Fisher, 1979.

The *source* of the feedback is recognized to influence an individual, with the associated construct of *source credibility* playing a significant role in the effectiveness of feedback. *Source credibility* refers to the perceived expertise and trustworthiness of the source of the feedback message. For feedback delivered by a clinical supervisor, *source credibility* represents a recipient's belief that the supervisor understands his tasks and work environment. If a healthcare provider perceives that his supervisor has a poor understanding of the required competencies of the work, or that the supervisor is not genuinely concerned about performance improvement, the feedback message is likely to be rejected. *Source credibility* may also be compromised by lack of trust in the data sources of the feedback message. For example, if a provider perceives that the quality of EMR data used to generate the feedback is poor, he may reject the feedback message.

The feedback *message* is described as information about past behavior. The properties of the feedback message include its information value as a measure of increase in knowledge or reduction in uncertainty about competing explanations for behavior. The message may also motivate future effort as an indicator of future rewards, or serve in as a kind of reward or punishment on its own to reinforce a behavior.

Perceived feedback primarily concerns how accurately the recipient perceives the mes-

sage that the source intended to provide. For example, a clinical supervisor who provides feedback may raise self-presentational concerns within a recipient that distracts him from accurately perceiving the feedback message. Peers who provide feedback on the other hand, and who may be more similar to the recipient, may provide feedback messages that are more accurately perceived. Ilgen and colleagues suggest that credibility of the source and power dynamics may influence the accuracy with which feedback is perceived. The authors also note that attributes of the message, including the sign and timing of the feedback, and the characteristics of the individual might influence how accurately the recipient perceives the intended message.

Acceptance of feedback is the degree to which the feedback message is accepted as an accurate and valid representation of her performance. This variable includes the perceived errors, fairness, and any negative affective reactions (e.g. discouragement, anxiety) to the feedback message that may cause it to be rejected. Both Kinicki et al. and Ilgen et al. claimed that negative messages are most likely to be misperceived and not accepted.

Next, *desire to respond* refers to motivational factors including external incentives and the recipient's intrinsic motivation. Desire to respond is determined by a) the ability to judge one's personal performance and b) locus of control, or the degree to which one believes she can freely choose to take action when performing tasks. *Desire to respond* follows *acceptance* because it is capable of reducing the impact of the message even if is accurately perceived and accepted. For example, even if a feedback message had a high perceived accuracy, for a healthcare provider who has devolved into learned helplessness, the condition under which one believes that locus of control is entirely external, he or she would not be motivated to respond to feedback.

The final moderating factor is *intended response*, which refers to the level of effort the recipient intends to dedicate to the task addressed by the feedback, relative to effort committed to other tasks competing for limited attention. The intended response is capable of moderating the effect of feedback in spite of a high desire to respond in cases where competition from other tasks of equal or higher importance cause a recipient to reduce the level of effort to the task addressed by the feedback.^{140,147}

There are several aspects of the FPM that are relevant to AF in health care at a fun-

damental level. Relative to FSC theories, FPM models the mechanisms that may influence an individual's approach to eliminating the FSC gap. For example, instead of increasing effort as a result of perceiving a feedback-standard gap, one might abandon the standard because of problems with source credibility or perceived accuracy of the feedback message. Changing or abandoning the standard could also occur when the perceived accuracy is adequate, but the recipient's desire to respond or intended response is low. Increasing effort, and resulting improved performance, follows an increased desire to respond and increased intended response. Therefore, FPM offers explanation for the decisions behind the resulting approaches to feedback-standard gap elimination that FSC theories describe.

Furthermore, FPM describes the rationale behind responses to feedback outlined by FSC theories, but specifically within the context of a workplace feedback intervention. One area of potential incompatibility between FPM and FSC theories is in Ilgen et al.'s assertions that negative feedback is likely to be misperceived and rejected. On the contrary, FSC theories require negative feedback as essential for understanding the feedback-standard gap. This discrepancy could be explained in part because of the broad scope of FSC theories, including specific and dissimilar tasks such as driving a car, team-based monitoring of radar screens, and playing tennis, where negative feedback can simply refer to any state other than the goal state. On the other hand FPM has been developed for the narrower scope of workplace performance feedback, where negative feedback represents a potential threat to the recipient's self-concept, and may refer to a summary of all performance, rather than performance for a specific task. Belief in the ineffectiveness of negative feedback in workplace feedback appears to be common within the field of industrial and organizational psychology, although some have argued for its use and effectiveness.^{108,148}

FPM may serve as a useful conceptual model to guide research about factors affecting the impact of feedback on performance in clinical settings. In applying FPM to clinical AF approaches, the theory emphasizes the importance of credibility of the feedback source, the perceived accuracy of the feedback, and, particularly relevant for AF in low-resource settings, the recognition that motivation to respond to feedback can be attenuated by intent to respond, due to competing priorities and limitations in the workplace. FPM is unique in providing a process model that emphasizes the impact of individual differences in feedback

recipients and their situations as moderators of the effect of feedback on performance. FPM aligns with the scenarios presented to illustrate COM-B, indicating that individual barriers to behavior change can differ, and that supervisors tailor feedback to accommodate such differences. Although FPM was developed as an explanatory model, it provides a potentially useful organizing framework for predicting the effect of feedback, as Kinicki et al. have demonstrated.¹⁴⁷ Therefore FPM could potentially be used to guide the tailoring of feedback messages according to observable differences in the cognitive variables it describes, and known individual and situational differences that influence each variable.

8.3 FEEDBACK INTERVENTION THEORY

Feedback Intervention Theory (FIT), proposed by Kluger and DeNisi in 1996, applies FSC theories to the context of a feedback intervention in a similar manner to FPM as a cognitive model of performance feedback processing. FIT additionally rests on the assumption that individuals maintain multiple goals that are hierarchically organized. At the top of the goal hierarchy are self-related goals (e.g. making a good impression) that control meta-task processes. This level of the hierarchy is referred to as the *self*. In the middle level are task-related goals (e.g. dedicating effort to a task) that control task-motivation processes, referred to as *focal task*, and at the bottom of the hierarchy are task detail-related goals (e.g. understanding how to perform the task) that control task-learning processes, referred to as *task details*. FIT also assumes a) that an individual's attention is limited and b) that feedback interventions strongly influence one's attention because of their potential to affect self-related goals. FIT proposes that feedback interventions which direct one's attention towards task-motivation (e.g. "You performed this task correctly 25 out of 30 times.") and task-learning processes (e.g. "NVP-containing regimens may cause jaundice, therefore discontinue any NVP-containing regimens for patients presenting with jaundice.") in middle and lower sections of the hierarchy are more likely to positively impact performance than feedback directs one's attention towards self-related goals and meta-task processes (e.g. "Your performance was excellent." or "Your performance is below the 50th percentile among your

peers.”).^{108,149}

FIT is in agreement with FSC theories, which claim that to affect performance, feedback should be specific and directly address the distance between performance and the goal. For example, in a feedback intervention designed to improve clinical performance, feedback about prescribing behavior that directs attention to the *task* level would inform a recipient about the level of adherence to the recommendation (e.g. 80% adherence) or her relative change in prescribing behavior compared to past performance. Feedback directing attention to the *self* level would inform the recipient about her prescribing behavior relative to peers, which does not directly address the gap between current performance and the adherence goal, and may elicit an affective reaction (e.g. happiness, discouragement) related to self-goals (e.g. “I perform better/worse than my peers”).

To test the notion that feedback can impact performance either negatively or positively depending on the attentional focus of the feedback message, Kluger and DeNisi conducted a meta-analysis of feedback intervention studies. Their meta-analysis included 131 studies and 607 effect sizes. They found that feedback had a moderate, positive effect on performance, with a weighted mean (adjusting for number of study participants) effect size of 0.41 (Cohen’s *d*). Although most interventions improved performance, feedback interventions worsened performance in more than 38% of the measured effects, demonstrating the inaccuracy of widely held assumptions that feedback uniformly improves performance.

To identify potential sources of negative effect of feedback on performance, the authors further conducted moderator analyses targeting factors that shift attention along the hierarchy of goals, either towards the self or towards the task. Kluger and DeNisi identified the following moderators of feedback effectiveness: feedback features (also called cues), task characteristics, and situational and personality variables. They were unable to measure the effect of personality on performance using a meta-analysis, but assert that there is strong support for the notion that personality variables moderate the effect of feedback, citing FPM as a primary source of support for this notion. The authors were able to conduct moderator analyses for all of the remaining moderators. They found strong support for the proposition that feedback features moderate the effect of feedback on performance according to their influence on attentional focus. For example, feedback interventions including praise

and verbal feedback were associated with lower performance, which was attributed to the fact that meeting in person raises self-presentational concerns. On the other hand, written, computer-based, and velocity feedback were associated with higher performance. The authors attributed velocity feedback to increased performance because it reveals the change in performance over time for the feedback-standard discrepancy. They also found weaker support for moderating effects from task characteristics and situational variables. Perhaps most significantly, their review demonstrated a non-significant effect of feedback sign on performance, suggesting that negative feedback can be used to improve performance.

FIT thereby demonstrates that performance feedback is a double-edged sword, having inconsistent effects on performance, based upon a wealth of evidence within the psychology literature.^{149,150} According to FIT, the features that are likely to increase feedback's impact are velocity feedback (showing change in performance), non-verbal feedback (delivered in writing or by computer), correct-solution feedback (providing information about correct actions to take), and feedback interventions that recommend or request an explicit performance goal. Features that are likely to diminish feedback's effect on performance include feedback that is provided verbally or in person, feedback that is threatening to self-esteem, and feedback that includes praise.

These findings are notably somewhat in contradiction to the 2012 Cochrane review of AF¹⁰, which found that feedback was more effective when provided by a supervisor or peer, whereas Kluger and DeNisi found that feedback provided in person is less likely to be effective, relative to feedback presented in writing. The difference in these findings may be attributable to differences in the wide range of environments that feedback was provided in for Kluger and DeNisi's review, relative to the narrower, clinical setting of AF interventions.

In contrast to FPM, FIT does not speak directly to cognitive variables such as *source credibility*, *perceived accuracy*, *desire to respond* and *intended response*. In this way FPM and FIT are complementary, emphasizing the influence of different factors on feedback efficacy. Unlike FPM, FIT claims that negative feedback can improve performance in a similar manner to positive feedback, whereas FPM implies more strongly that negative feedback is likely to be rejected and should be excluded from a feedback intervention. FIT could best be related to FPM as adding a new dimension of attentional focus to the analysis of factors affecting

feedback efficacy, with an emphasis on the attentional focus of feedback features.

In relation to FSC theories, FIT clearly aligns with the notion that feedback should address the feedback-standard gap, in that when attention is directed to the focal-task and task-learning levels, feedback is going to more directly describe the feedback-standard gap. Conversely, when feedback directs attention towards the self, the feedback-standard gap is likely to be described in relation to the goals of the self, bringing tangential information into the feedback message that can weaken its impact on performance.

FIT can significantly contribute to the interpretation of AF evidence, as demonstrated by Hysong et al.'s meta-analysis. FIT potentially explains the variability in studies of AF that show both increases and decreases in performance resulting from a feedback intervention, in conjunction with the attentional focus of the feedback. FIT may not support the use of individual, in-person meetings with a supervisor to deliver verbal feedback because of the likelihood of self-presentational concerns interfering with perception of the feedback-standard gap. Instead, FIT suggests that individual feedback delivered in writing would have a greater effect than the same feedback delivered verbally. Similarly, FIT aligns with FSC theories in suggesting that social comparison information would not improve the impact of feedback on performance because social comparison directs attention to the goals of the self, rather than to the task details.

Like FSC, FIT appears to align completely with the feedback features identified by Hysong's actionable feedback model. Timeliness, individualization, non-punitiveness, and customizability are all features of feedback that enable the attention of the recipient to focus on task-details and the feedback-standard gap. In particular, non-punitiveness reflects the exclusion of threats or punishment that would direct attention to the self. Customizability could be said to be used to identify self-related goals, if, for example, customization enabled social comparison feedback, which might permit feedback to direct attention to the self rather than the task, but this is dependent upon the level of customization - individual, clinic, or hospital customization and, in the case of individuals, the goals that are set. For example, if an individual has a goal of achieving performance within the top 10% of peers, this is a goal that requires feedback to direct attention to self-related goals.

Hysong's meta-analysis of trials of AF reveals a dimension of FIT that is problematic

given the degree of under-reporting of feedback features for the context of evidence synthesis. FIT posits that graphical feedback would direct attention to task details more than to the self, but this effect is justified according to the same terms as written feedback in that, relative to verbal feedback delivery, written and graphical feedback do not require an individual to interact with a supervisor individually, which could raise concerns about the self. This claim is conditionally relevant to feedback provided verbally, but is not sufficient to claim that graphical feedback directs attention towards the task and away from the self. This is because graphical presentation of feedback can be used to display information that directs attention to the self, such as social comparison feedback, (described as normative information by Kluger et al.). Hysong et al. mention the fact that social comparison feedback, showing change in peer performance over time, is better-suited to graphical presentation, which lends support to their finding that graphical feedback decreased performance, contrary to their expectations set by FIT. Therefore, synthesizing evidence based on the principles of FIT creates the possibility that non-reported factors, such as the provision of graphical feedback containing normative information, can confound the resulting effect size ascribed to the theory. This is the same problem that Gardner et al. encountered when conducting a meta-analysis of AF studies using control theory as a framework for predicting the successful use of AF, in which under-reporting of the study prevented the authors from determining if the intervention was fully aligned with the principles of control theory. These theories are likely to yield stronger results when used prospectively to inform intervention design, rather than when used for evidence synthesis due the multi-faceted nature of performance improvement interventions, and the extent to which aspects of the study that reveal theoretical misalignment are under-reported.

FIT's primary implication for AF interventions is that the feedback message may worsen performance when it directs attention to the self. This suggests that feedback messages might be improved by avoiding the use of peer comparison information and information conveying judgment. Interventions may also be more effective when feedback is limited to presenting information about the recipient's past performance or about how to perform a task. FIT also supports the notion that individual differences moderate the effect of feedback on performance, although this is not a central finding of the meta-analysis conducted by the

authors.

8.4 REGULATORY FOCUS THEORY

Regulatory Focus Theory (RFT), like FPM and FIT, adds another dimension of feedback's influence on performance to the fundamental understanding provided by FSC theories. Proposed by Higgins in 1997, RFT is based on the hedonic principle, which states that people seek to avoid pain and to pursue pleasure. According to RFT, two concurrent regulatory systems within each person influence the ways in which one avoids pain and approaches pleasure. These two systems operate continually and competitively, responding to multiple influences that activate either system into predominance. Each system is a source of motivation to reach a desired end state: either the avoidance of pain or the approach of pleasure. Higgins describes these two systems as *prevention focus* (avoidance of pain or loss/non-gain) and *promotion focus* (approach of pleasure or non-loss/gain). Higgins describes an individual under promotion focus as having a greater sensitivity towards positive outcomes (non-loss/gain), where prevention focus represents one having a greater sensitivity towards negative outcomes (non-loss/gain). Under prevention focus, an individual is concerned that something bad that might happen, or has something to lose and nothing to gain. Prevention focus influences individuals to prioritize adherence to rules, accuracy, and safety. Actions taken under prevention focus feel like obligations or requirements. Under promotion focus, an individual aims to achieve something new, and feels there is nothing to lose but something to gain, therefore he or she will prioritize creativity, open-mindedness, and growth. Individuals taking actions under promotion focus feel that they are pursuing their desires. The resulting affective reactions to achieving or not achieving the desired end states are also aligned with the two self-regulatory systems. Under prevention focus, if an individual perceives avoidance of some negative outcome, he will feel relaxation. However if the opposite is perceived, that some negative outcome is not being avoided, he will feel anxiety. Under promotion focus, if an individual perceives approach of a positive outcome, he will feel happiness. However if the opposite is perceived, he will feel discouragement.

Happiness and anxiety are high-arousal emotions that indicate high motivation, while discouragement and relaxation are low-arousal emotions that indicate low-motivation. Thus as feedback indicates avoidance of negative outcomes or approach of positive outcomes, high and low motivation can result from either positive or negative feedback, as a function of the predominant regulatory focus. In other words, feedback, whether positive or negative, can cause either an increase or a decrease in motivation, depending on regulatory focus (Figure 9).^{151,152}

		Regulatory Focus	
		Prevention	Promotion
Feedback Sign	Positive	<p>Decreased motivation</p> <p>Affect: Relaxation</p>	<p>Increased motivation</p> <p>Affect: Happiness</p>
	Negative	<p>Increased motivation</p> <p>Affect: Anxiety</p>	<p>Decreased motivation</p> <p>Affect: Discouragement</p>

Figure 9: Regulatory Focus and Feedback Sign

Multiple factors can combine to activate one regulatory focus into predominance, or to balance between the two foci. These factors include the type of outcome one is working towards, aspects of the task, the context in which the task is performed, and the individual’s personality. Examples of outcomes that may emphasize prevention focus (loss/non-gain) in health care that may influence regulatory focus are working to help a patient recover under critical care, or identifying adverse drug events. An example of a promotion focus (gain/non-loss), outcome is implementing workflow changes to improve efficiency in a clinic, or designing interventions to improve patient satisfaction. Tasks that would inherently emphasize prevention focus involve error detection, such as when reading a radiological scan for cancer screening, or following rules within a clinical protocol. On the other hand, an

example of a promotion task in health care would be creative work that requires one to generate ideas, such as creating a differential diagnosis, or designing a clinical trial. In addition to the task type, according to Higgins, task instructions can convey gain-nongain or nonloss-loss information that influence the promotion or prevention focus of an individual when performing the task. Task instructions may influence regulatory focus by including incentives, such as pay-for-performance, or threats in the form of penalties. The context in which a task is performed also influences one's regulatory focus. For example, if a person has taken a job because she had no other employment opportunity, her context would encourage a prevention focus to a greater degree than someone who feels that a job is contributing to the pursuit of her ideal career. A final significant influence on regulatory focus stems from an individual's personality, called her chronic regulatory focus. Chronic regulatory focus describes a person's general need for security relative to her need for growth. An individual's chronic regulatory focus may be strongly oriented towards one or the other focus, or may be somewhere between the two and could result in a neutral chronic focus in which an individual's needs do not significantly influence her regulatory focus. Similarly, the influence of any factor is not necessarily bi-modal, and may not clearly contribute to the activation of one system or another. The degree of alignment between the influences of outcome, task, context, and personality on regulatory focus is known as the *regulatory fit*.¹⁵⁰

Regulatory fit is described by Higgins as degrees of match and mismatch between the influences of regulatory focus, such as a person's chronic focus, the properties of the task and the context in which it is performed. If the regulatory fit is good, the perceived value of the task increases, which is likely to translate into increased motivation and performance. Studies by Van Dijk and Kluger have demonstrated the interaction of regulatory fit, feedback sign and their impact on motivation. In a 2004 study they assessed the chronic regulatory focus of study participants and then evaluated changes in motivation in response to simulated performance feedback, finding that provision of feedback improved motivation with good regulatory fit, and decreased motivation under poor regulatory fit.¹⁵³ A 2010 study demonstrated that positive feedback improved motivation and performance for promotion-focus tasks requiring participants to generate ideas, compared to performance of prevention-focus tasks that required participants to detect errors. Conversely, positive feedback decreased mo-

tivation and performance for prevention-focus tasks.¹⁴⁸ Therefore, the combined influence of the task, individual personality, and context combine to establish a regulatory focus for each individual. When an individual's regulatory fit is good, meaning the influences all contribute to one regulatory focus, motivation will be highest. When regulatory focus is more strongly oriented towards promotion, with good regulatory fit, individuals will increase performance when receiving positive feedback, and decrease performance when receiving negative feedback. When regulatory focus is more strongly oriented towards prevention, individuals will decrease performance when receiving positive feedback, and increase performance when receiving negative feedback.

The primary implication of regulatory focus theory (RFT) for FSC theories are that regulatory focus mediates the motivation resulting from the knowledge of the distance between current performance and the goal. To some degree, RFT describes the differences between goal-setting theory and control theory, which emphasize promotion and prevention aspects of discrepancy reduction, respectively. RFT is relevant to FPM in its alignment with the *desire to respond* and *intended response* constructs. Despite providing feedback with high perceived accuracy and validity, the effect of feedback on performance may be attenuated by a decreased motivation as a result of poor regulatory fit. Where FPM identifies motivation as a cognitive variable (*desire to respond*), RFT endeavors to explain the mechanism by which this cognitive variable operates. RFT and FIT appear to address different mechanisms of the impact of feedback on performance. FIT postulates that feedback directing attention to the self will have a reduced impact on performance, while RFT suggests that feedback provided under different regulatory foci can have divergent effects on motivation and performance. Therefore, by contributing a greater understanding of the function and influences on motivation, RFT is complementary to FSC, FPM, and FIT.

RFT provides important considerations for AF interventions. In clinical settings using critical incident AF, in which providers aim to learn from a medical error that is often attributed to a breakdown in the clinical system, the focus on a negative outcome, patient safety, and routine care seems very likely to create a strong situational prevention focus. For example, prevention focus could be activated by the highly significant avoidance goal of preventing maternal death or hospital-acquired infection, which are events that all clin-

icians are likely to unambiguously agree to try to avoid. Critical incident AF, which is outcome-focused, is only held when negative outcomes occur, setting up an alignment with the motivational mechanisms proposed by RFT. For example, when no negative incidents occur, there is no review meeting called to discuss the positive nature of the clinic's performance, which would elicit feelings of relaxation and may lead to decreased performance. The alignment of regulatory focus and feedback sign suggests that AF is likely to succeed when used for preventing negative clinical outcomes.

Guideline implementation feedback on the other hand may not benefit from the same alignment of regulatory focus and feedback sign. Routine guideline adherence feedback may deliver positive feedback to clinicians under prevention focus, effectively resulting in feelings of relaxation, which according to RFT is likely to decrease performance. To avoid decreasing performance, guideline implementers might consider withholding positive feedback about avoidance goals that are likely to emphasize prevention focus. However, providing frequent, negative feedback increases the risk of causing abandonment or changing of the performance standard, which is an equally undesired outcome in the context of guideline implementation interventions.

Given the impact of regulatory fit and the resulting regulatory focus on motivation in response to feedback demonstrated by Van Dijk and Kluger, RFT represents an important area of investigation for research evaluating AF interventions. Watling and colleagues explored the application of RFT in a retrospective analysis of a qualitative study using grounded theory to interview early-career academic doctors about their experiences in clinical learning.¹⁵⁴ The authors found that RFT could provide insight in situations where one focus or the other was clear in some situations. However, many situations were difficult to interpret using RFT because of the tendency for multiple competing influences to create a balanced regulatory focus and because of the possibility that regulatory focus can change over time. Van Dijk and Kluger have supported the notion that understanding an individual's regulatory focus is difficult, and include it as a reason to avoid feedback interventions in medical training contexts.¹⁵⁰

While RFT may underscore the difficulty of getting feedback right in clinical settings, this theory also supports the notion that feedback should be tailored for individual differences.

It appears that no AF intervention, to our knowledge, has demonstrated a useful approach for providing individualized feedback using RFT. However, the theory nevertheless strongly supports the notion that feedback interventions are likely to be more effective if they can accommodate individual differences.

8.5 GROUP FEEDBACK PROCESSING

Unlike research addressing individual task performance, the body of group feedback research is smaller, but offers insights about the potential mechanisms by which group feedback affects group and individual performance.¹⁵⁵ Research investigating group feedback addresses heterogeneous types of feedback, tasks, groups, and work contexts. A common definition of “group”, synonymous with “team”, is defined by Salas as “a set of two or more people who interact dynamically, interdependently and adaptively toward a common and valued goal, each having specific roles or functions to perform and a limited life-span of membership.”¹⁵⁶ Researchers have studied group feedback processes and the effects of feedback on group performance within laboratory and field settings. Important aspects of group task performance that may moderate the effect of feedback include task interdependence, individual or collective group orientation, and group feedback processing.

Task interdependence refers to the degree to which group members must coordinate actions to successfully perform a task. As task interdependence increases, task responsibility is increasingly shifted from individuals to the entire group. To successfully perform interdependent tasks and achieve goals, groups require different levels of coordination. The amount of coordination required to perform interdependent tasks successfully is a function of the degree of conflict between group goals and individual goals. Conversely, the degree to which individual and group goals do not conflict and are additive, (i.e. individual task performance always contributes to both individual goals and group goals) is indicative the amount of coordination required for successful task performance, where completely additive tasks require no coordination. When individual goals conflict with group goals, individuals must regulate their behavior in order to balance between pursuit of multiple goals. For example,

within ART clinics in Malawi, due to staff shortages, conflicting individual and group goals are held by clinical officers (COs). COs are frequently assigned to cover additional hospital wards on the same day as they are required to work in the ART clinic. COs must regulate their work activity to balance the demand for patient care in multiple clinics in order to meet their individual goals and to meet the ART clinic's group goal. Therefore group work is characterized by interdependent tasks that require some level of coordination in order to prevent individuals from working towards individual goals at the expense of group goals.¹⁵⁶

In recognizing differences between group and individual task performance, it follows that customizing feedback for either groups or individuals may improve its effect on performance for either type of task. Research indicates that group feedback is more appropriate for interdependent, group tasks, while individual feedback is more appropriate for individual tasks.¹⁵⁷ A 2004 study comparing the effect of group, individual, and combined feedback on group task performance demonstrated that feedback can significantly influence an individual's decision to prioritize individual or group goals, affecting the group's ability to achieve its goals.¹⁵⁶ If the provision of group or individual feedback can influence the way that an individual balances conflicting goals, then an important question becomes "When is group or individual feedback more relevant to improving performance for a given task?" According to Deshon, the value of group feedback increases with the amount of coordination required to prevent conflict between individual and group goals. Therefore, for interdependent tasks that require significant coordination, group feedback is more appropriate, while for individual tasks and interdependent tasks that require little coordination, individual feedback may be more appropriate.

Another important factor affecting responses to group feedback is a group's individual or collectivist orientation. When groups receive group feedback, it fosters a collective orientation among group members in addition to influencing the prioritization of group goals. Similarly, providing group members with individual feedback can foster individual orientation and increase prioritization of individual goals. This may have important influences on group performance in light of the group's ability of the group to meet its goals.^{158,159} Individual differences in individual or collectivist orientation can significantly moderate the effect of individual or group feedback on performance, according to a review of group feedback

research.¹⁵⁷ Furthermore, within groups, power imbalances can interact with individual or collective orientation affect the level of group performance. A 2010 study examined group power imbalances and the degree to which power imbalances can be leveraged to degrade or improve group performance. The study found that the degree of individual or collective orientation of the group members became a significant factor associated with the use of power imbalances to improve or degrade group performance. Where a group had a collectivist orientation, power imbalances improved group performance because more powerful members of the group leveraged their power to improve performance for the entire group. The opposite effect was observed in groups with an individualist orientation, because more powerful members prioritized their own performance at the expense of achieving the goals of the entire group.¹⁵⁹

A final primary difference between individuals and groups occurs in feedback processing. Individual and group responses to feedback are similar in that both have tendencies to attribute success to the group or individual when receiving positive feedback, and both are likely to blame the environment or external factors in the case of negative feedback. However when processing feedback, groups seem to be able to use feedback for learning more effectively than individuals, and groups process feedback with less variability than individuals, making it more likely for groups to respond more consistently to feedback.¹⁵⁸

An important implication of research addressing the provision of group feedback for AF interventions is that individualized feedback should only be provided when the task being performed does not require significant team coordination. Although this may seem rather simple as an implication, it represents an important criterion for a successful feedback intervention.

Another implication of group feedback research for AF interventions that is a collectivist orientation among feedback intervention stakeholders could lead to the discouragement of or a devaluing of provision of individual feedback, both on the part of supervisors and providers receiving feedback, relative to stakeholders in an individualistic culture.

8.6 SUMMARY

I reviewed the literature on theories of feedback primarily relating to the provision of performance feedback in organizations. Because of the enormity of literature addressing performance feedback, I limited the literature I reviewed to only that which appeared most relevant to AF interventions in clinical settings, and which provided the largest syntheses of existing work. It is possible that this review does not include theories that hold important implications for the design of clinical AF interventions.

Decades of research into the effectiveness of feedback interventions in organizations have yielded insights that may lead to improvement of the design of AF interventions, even though the causal mechanisms by which feedback impacts performance are not well understood. Despite a long history of development of feedback theory, the literature remains fractured across theories that are mostly complimentary. Feedback theories fundamentally rest on the feedback-standard comparison, which claims that motivation arises from a desire to eliminate the perceived discrepancy between a current level of performance and a goal or standard. FPM expands upon FSC theories by identifying cognitive variables that mediate the effect of feedback on performance and affect perception of the feedback-standard discrepancy. FIT contributes to FSC and FPM by accounting for multiple-goal contexts and evaluating the impact of feedback's attentional focus as a further mediator of feedback's effect on performance. RFT contributes understanding of the interaction between motivation and feedback sign, and of the complex nature of contextual factors as moderators of feedback effectiveness. Research on group feedback indicates that task coordination levels can be used to understand when group feedback is likely to be more effective than individual feedback for improving group performance.

Feedback theories provide valuable insights for feedback interventions within healthcare, but the application of feedback theory to AF interventions has, until recently, been rare. Beyond FSC theories that describe more fundamental mechanisms of feedback interventions, FPM, FIT, and RFT provide strong support for the notion that the effectiveness of feedback is dependent on a highly contextualized and complex set of factors, many of which are directly attributable to the individual feedback recipient.

9.0 TAILORING CLINICAL PERFORMANCE FEEDBACK

A primary purpose of AF interventions is to change clinical behavior. Theories of behavior change and theories of performance feedback align to strongly support the idea that AF interventions are likely to be more effective when they provide individualized feedback that is relevant to individual barriers to behavior change and individual characteristics affecting feedback processing. The term “tailoring” is used variably in the literature to describe processes of adapting a behavior change intervention for an individual or population.¹⁶⁰ In the implementation science literature, “tailoring” has been used to describe the process of mapping an intervention to barriers and facilitators of knowledge use within an intervention population.^{161,162} I use the term “tailoring” in the implementation science sense, referring to a process of prospectively identifying barriers to behavior change, and then changing some features of the intervention to match identified barriers. I use the concept of *automated tailoring* to refer to computer-assisted planning of an intervention in a dynamic and continuous process that addresses both shared and individual barriers and facilitators for the duration of the intervention. Related research in the fields of biomedical informatics and health communication provide established methods that, to my knowledge, have not been applied for the purpose of tailoring feedback messages within AF interventions.

9.1 FEEDBACK TAILORING IN MEDICAL KNOWLEDGE-BASED SYSTEMS

In the field of biomedical informatics, mitigating the complexity of the clinical environment and clinical cognition to provide relevant feedback messages is a primary goal of med-

ical knowledge-based systems that have been developed and refined over nearly the last half-century.¹⁶³ Informatics researchers and developers of medical knowledge-based systems such as clinical decision support systems,¹⁶⁴ intelligent tutoring systems,¹⁶⁵ and computer-interpretable clinical guidelines¹⁶⁶ use knowledge representation methods to identify opportunities to provide computer-generated feedback.

From an informatics perspective, knowledge is defined as “relationships, facts, assumptions, heuristics and models derived from the formal or informal analysis (or interpretation) of data.”¹⁶³ Knowledge representation is a sub-field of artificial intelligence that has received significant research attention in biomedical informatics. Knowledge representation research concerns the explicit encoding of knowledge artifacts to enable a computer program to reason with data using inference methods and problem-solving strategies.

Knowledge-based systems contain a knowledge base that operationalizes knowledge artifacts for clinical purposes. Medical knowledge-based systems typically use a qualitative model of the relationships between inferences to draw abstract conclusions about a patient or situation, which can include probabilistic, causal and temporal relationships.¹⁶³ These models are often referred to as ontologies.

9.1.1 Ontologies and knowledge representation artifacts

Within knowledge-based systems, knowledge representation artifacts have been variably described as ontologies, including artifacts that represent terms, concepts, and real entities.¹⁶⁷ A commonly-cited definition for the term ontology is “a specification of a conceptualization.”¹⁶⁸, but this definition does little to resolve ambiguity about its meaning, which has been debated.¹⁶⁹ Schulz and Jansen discuss formal ontology and its relationship to other knowledge representation artifacts in the context of biomedicine.¹⁶⁷ They describe knowledge representation artifacts according to four types of statements that I find helpful for conceptualizing artifacts that are used within a knowledge base. These statement types are *universal*, *terminological*, *assertional*, and *contingent*.

Universal statements refer to entities within an objective reality (from a logical positivist stance) that are not context-dependent and that are the most stable of descriptions. These

statements are represented using a description logic, such as Web Ontology Language (OWL), and supported by software tools like the Protégé editor. OWL uses classes, class properties, instances (also called class members) and the relationships between classes to create decidable statements, or those which can be mathematically evaluated to enable a computer to reason with data.¹⁶⁷ An example of such an ontology is the Foundational Model of Anatomy (FMA), which contains encyclopedic knowledge of human anatomy.¹⁷⁰

Terminological statements are contained in controlled vocabularies and thesauri, referring to natural language used to describe entities. These statements are used to support mapping across terms for such purposes as natural language processing, system interoperability, and information retrieval. Examples of terminologies include controlled vocabularies such as Medical Subject Headings (MeSH), Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT), and Logical Observation Identifiers Names and Codes (LOINC).¹⁷¹

Assertional statements describe individuals and class instances, whose properties may be represented within a formal ontology. These are typically contained in a relational database, for example in a medical record system that contains observational data about patients.

Finally, contingent statements describe attributes of a class that may or may not be true for all members, and hence are not universals. These knowledge artifacts typically include probabilities, decision rules, and causal relationships that are context-sensitive.

Taking this view, knowledge-bases are comprised of artifacts including a formal ontology (containing universal statements) that is an organizing foundation for other knowledge representation artifacts. These other artifacts include terminologies (containing terminological statements), relational databases (containing assertional statements) and decision rules or probabilistic models (containing contingent statements). In reality, a clean separation of these statement categories is not always feasible and may not even be desirable. However, the approach Schulz and Jansen describe may help knowledge-based system developers to avoid semantic ambiguity that leads to confusion about the use of terms and knowledge-base maintenance challenges.¹⁶⁷

9.1.2 Probabilistic knowledge representations

Probabilistic knowledge representation methods are used in biomedicine to reason with data about biomedical knowledge under conditions of uncertainty. Probabilistic knowledge representation methods have been used in biomedicine for diverse applications such as decision support tools in medical diagnosis and prognosis, intelligent tutoring systems, and disease outbreak surveillance. An area of growing interest in biomedical informatics in which probabilistic knowledge representation may play a significant role is the development of “high-throughput phenotyping” of electronic health record (EHR) data.¹⁷² This work involves the study of EHRs as natural systems whose features are analyzed using machine learning and data mining approaches. Phenotyping of EHR data aims to understand the meaning of features in the clinical data on a large scale despite challenges presented by missing data, complexity in care processes and errors that create inaccurate data. Hripcsak and Albers have suggested for example, that by analyzing raw EHR data, a probability of disease could be reliably predicted, rather than relying on a query of the clinical data that requires significant effort to develop and maintain.

In a process similar to EHR data phenotyping, analysis of the features of clinical performance and associated eHealth data could enable the reliable calculation of a probability of the determinants of clinical behaviors such as the capability, opportunity, and motivation of healthcare providers who use eHealth, which could then be used to inform the tailoring of clinical performance feedback. Probabilistic models developed in human-computer interaction research have been used to predict the cognitive state of computer system users’ emotions¹⁷³, self-efficacy¹⁷⁴, and attention.¹⁷⁵ The growing use of these methods suggest that their application to the analysis of clinical data could support the automated tailoring of performance feedback.

9.1.3 Tailoring of clinical feedback

To my knowledge, the tailoring processes that occur within biomedical knowledge-based systems are largely implicit, resulting from the tendency to provide immediate or near-immediate feedback that is highly relevant to the present situation of the recipient. Because

such systems as clinical decision support systems and intelligent tutoring systems observe and infer behavioral and cognitive activity in real-time or near real-time, the computer-generated feedback they provide is situationally tailored. Knowledge-based systems that provide feedback can use production rules or probabilistic models to anticipate the meaning of the recipient's actions in the present situation. Situational tailoring is a function of the system identifying opportunities to impact the recipient's awareness in the present situation.

Unlike the provision of immediate feedback which is effectively situationally tailored, the provision of summary feedback may hold different implications for message tailoring. Summary feedback provided in an AF intervention encompasses behavior enacted over many situations in a complex and dynamic clinical environment. Therefore tailored summary feedback must account for a range of situations rather than a single situation, which raises the possibility that a summary feedback message may be more relevant to some situations than others within a reporting period. While the automated tailoring of summary feedback for healthcare professional behavior change has not been studied in the field of biomedical informatics to my knowledge, related work in the field of health communication provides evidence and methods that may yield insight into how to design persuasive automated feedback message tailoring systems for AF.

9.2 COMPUTER-TAILORED HEALTH COMMUNICATION

Health communication is an inter-disciplinary field at the intersection of public health, communication studies, psychology, medicine and community development that concerns the influence of communication on health-related behaviors and decision making. Health communication has been defined as “the art and technique of informing, influencing, and motivating individual, institutional, and public audiences about important health issues.”¹⁷⁶ Over the last two decades, researchers have developed theoretical foundations and developed approaches for computer-tailored health communication (CTHC), which now commonly involves algorithmic tailoring of messages for individuals by a software application.¹⁷⁷

CTHC methods have used psychological theory to dynamically adapt feedback messages

over time, targeting behaviors such as physical exercise, smoking cessation and nutritional intake.¹⁷⁸ CTHC messages aim to persuade individuals by addressing behavioral determinants that are usually identified using theory-informed questionnaires or data collected in clinical records.¹⁷⁹ Tailored health message interventions may not include performance feedback, instead delivering generalized information, for example, about the benefits of behavior change or the risks of continuing unhealthy behaviors.

The theoretical rationale for CTHC is based on the Elaboration Likelihood Model, which asserts that individuals' cognitive processing of messages is more likely to lead to behavior change when the message is more individually relevant.¹⁶⁰ The specific relevance of the message to the individual spans a range of characteristics such as reading ability, lifestyle preferences, cultural perspective, beliefs about learning, and present circumstances. These characteristics are matched with message components using an algorithm containing rules developed by behavioral experts. The algorithm is designed to bring only the minimal amount of information necessary to the individual in the most acceptable format.^{179,180}

CTHC researchers have recently explored approaches that leverage a recipient's message preference data to adaptively tailor messages.¹⁸¹ This work is based on recommender system approaches, such as matrix factorization, used by companies such as Netflix and Amazon that create personalized recommendations based on analysis of consumer feedback. Feedback from consumers can be explicit feedback, from data such as ratings (e.g. the number of stars an individual chooses to rate a product) or implicit feedback, from behavioral data such as purchasing activity, reading time spent, and page views.^{182,183}

Evidence about CTHC approaches has shown that automated, individualized message tailoring can significantly improve health behaviors.¹⁸⁴ Evidence indicates that CTHC approaches that dynamically adapt messages for individuals over time are more effective at changing behavior, and that tailored messages which address multiple behaviors simultaneously were as effective as messages that targeted a single behavior. Research in CTHC offers more than two decades of knowledge developed that is closely related to tailoring of automated performance feedback messages for AF. I anticipate that, while there are significant differences in the context, motivations, and behaviors addressed by CTHC and AF, there is much CTHC knowledge from an informatics perspective that can be directly applied to

automated message tailoring for AF.

9.2.1 AUTOMATED TAILORING OF HEALTH MESSAGES

A key example of the CTHC knowledge that may be useful for tailoring clinical performance feedback is the process of developing tailoring algorithms for the automated tailoring of feedback messages. Kreuter et al. describe the process of developing tailoring algorithms as a linkage of the data collected that is relevant to tailoring, the design templates for messages, and the content of the messages.¹⁷⁷ The tailoring algorithm links these components by operating on three types of variables: *raw variables*, *intermediate variables*, and *feedback variables*.

Raw variables represent the data collected from or about the message recipient, including questionnaire responses and medical record data. Questionnaire responses include reported information about health activities such as the number of cigarettes smoked per day, amount of physical exercise, and or answers to specific questions such as “In the last six months, has your doctor told you to get more physical activity?” Questionnaire responses may also include data about the recipient’s reported barriers to behavior change. *Raw variables* may also include medical record data such as vital signs, test results, personal identifiers, and demographic information that are abstracted with the permission of the recipient. The concept of *raw variables* can be used to describe to data collected about clinical behavior for message tailoring. These data could include the current performance data for an individual, such as the percentage of eligible ART patients that an received a prescription for CPT. *Raw variables* can also include the performance history for an individual, and that of their peers. Other relevant data may include an individuals’ reported preferences for receiving feedback, and their demographic data, including work experience and professional role.

Intermediate variables represent data that is transformed, summarized or calculated by a classification algorithm or other tailoring rules that process the raw data. An example of an intermediate variables for health message tailoring is a recipient’s BMI, calculated using the recipient’s current height and weight. Other examples include using classification, such as selecting the recipient’s age group based on current age, and calculating a cigarette addiction

level using the number of cigarettes smoked each day and the time of day when the first cigarette is smoked. The concept of intermediate variables can be applied to tailoring clinical performance feedback in terms of identifying the features of performance and situational factors. For example, performance could be classified as “low”, “moderate”, or “high” based on the current performance level, and performance of peers working in the same clinic during the reporting period could be similarly classified as a situational feature.

Finally, *feedback variables* represent the feedback message components that will be used in a tailored health message. These components are sentence fragments, numeric values, and terms that correspond with the individual’s context and barriers to behavior change. For example, the feedback message could refer to a supportive person that the recipient has indicated they will rely on to change their health-related behavior. The feedback message components for the supportive person could include “member of your weight-loss support group”, “doctor or other healthcare professional”, or “spouse, partner, or family member”. The appropriate phrase would be used according to the choice made by the individual when completing a questionnaire about the targeted behavior. The concept of *feedback variables* could be equated to feedback message components used in presenting performance feedback, including the use of peer comparison, performance history, and various graphical formats that could be selected.

These three types of variables are used by a tailoring algorithm to identify the relevant features of the individual and their situation, assess the meaning and implications of these features for the provision of feedback messages, and to select the message components that are most likely to improve behavior change. The tailoring algorithms described by Kreuter et al. use decision rules to process the raw variables, then the intermediate variables, and finally both the intermediate and raw variables are processed again to instantiate the feedback variables. The established use of these processes for health behavior change, supported by evidence showing their efficacy, suggests that they may be effective for tailoring of feedback messages for clinical audit and feedback.

9.3 SUMMARY

I have presented an overview of research relevant to the automated tailoring of performance feedback messages for healthcare professionals. A fundamental motivation for this research is the existence of persistent gaps between what is recommended by medical evidence, and the decisions that are routinely made by healthcare professionals, which have created a need for continuous learning in healthcare organizations. The growth of eHealth is both driving the increased pace of biomedical knowledge generation, and creating opportunities to automate the implementation of new knowledge within clinical settings. Contingencies regarding the quality of clinical data used for the purpose of performance measurement must be recognized and understood. These opportunities are pronounced in low-income countries where an eHealth infrastructure is already being established and supervision resources are limited.

Both AF evidence and relevant psychological theory suggest that clinical performance feedback is an important requirement for continuous learning and behavior change, but that we do not understand very well how or when feedback interventions work. Psychological theory can guide implementation interventions by providing insight into the causal mechanisms that enable AF interventions to be effective. Theories of behavior change and performance feedback suggest that tailoring messages for individual and situational differences could significantly improve interventions. The knowledge representation and message tailoring methods required to successfully build a feedback tailoring system have been under development for some time in the fields of biomedical informatics and health communication.

The opportunities, evidence, theories, and methods that I have discussed form a foundation on which tools could be developed that support the tailoring of clinical performance feedback. This work involves uncertainty both in the use of eHealth data for performance measurement and in the provision of effective feedback (Figure 1). The challenges of performance measurement primarily concerns questions about the fitness-for-use of clinical data, whereas the challenges of providing effective feedback primarily concerns our understanding of the causal mechanisms that feedback leverages to influence clinical learning and behavior change. These challenges are closely related however, as data quality and knowledge

representation are defined by the purpose of use, which is to support clinical learning and behavior change. Furthermore, as technology provides new opportunities to measure performance, new types of feedback can be generated that were not possible or affordable before the introduction of the technology.

My goal therefore is to design performance feedback tailoring tools that support the generation and testing of hypotheses about the effectiveness of feedback in clinical settings, about the use of eHealth data for performance feedback, and about the tools themselves that support both tasks. I aim to understand and apply these challenges in the context of a low-resource setting in which performance can feasibly be measured, where feedback is likely to be effective, and where the potential for implementation of an automated feedback message tailoring tool could extend to a national scale.

10.0 RESEARCH DESIGN AND METHODS

The primary goal of this research was to understand the interacting challenges of performance measurement and feedback effectiveness in a low-resource setting. I developed and applied this understanding in the context of ART clinics in Malawi's public hospitals where an EMR is used. This research involved a combination of methods that included theoretical modeling, information system design and formative evaluation, and qualitative and quantitative data analyses.

I began by asking where it was feasible to measure individual clinical performance using EMR data and recommendations from a clinical guideline document. Having identified a set of measures in which individual performance could be routinely measured, I studied the clinical environments and healthcare providers to learn about the potential contextual barriers that could prevent feedback from influencing clinical learning and behavior change. After identifying barriers, I designed a novel information system that could enable feedback to overcome or avoid these barriers by using psychological theories to guide individualized tailoring of performance feedback messages. Finally, I evaluated the design of the information system and its potential impact in Malawi by using psychological theories to guide the analysis of de-identified EMR data, to understand how frequently feedback messages could be tailored.

This process is described in four studies, guided by the following four research questions:

Measuring individual performance using eHealth data in low-resource settings:

Can we use EMR data and a guideline document to measure individual clinical performance in ART clinics in Malawi?

Barriers to using eHealth data for feedback in low-resource settings: What contex-

tual factors may prevent us from successfully using EMR data for automated audit and feedback in ART clinics in Malawi?

Designing knowledge-based clinical feedback message tailoring: How could audit and feedback continually adapt to changes in complex healthcare environments?

Automated feedback message tailoring in low-resource settings: What is the potential impact of automated feedback message tailoring on clinical performance in ART clinics in Malawi?

11.0 MEASURING INDIVIDUAL CLINICAL PERFORMANCE USING E-HEALTH DATA IN LOW-RESOURCE SETTINGS

The objective of this study is to understand the feasibility of measuring individual performance of healthcare providers using existing EMR data and a published national guideline for delivery of antiretroviral therapy (ART) in Malawi, Africa, for the purpose of generating automated performance feedback. I also sought to characterize and measure the factors that limit the provision of automated feedback in this setting. This research was published in JAMIA in 2011.¹⁸⁵

11.1 METHODS

My goal was to understand if we could create guideline-based performance indicators that an automated system could use to measure individual clinical performance for ART clinics in Malawi. To achieve this goal, I created an evaluation process to identify auditable guideline recommendations using a published guideline document and available EMR data. For this evaluation I used only CPG statements that directly addressed clinical management of adult ART patients from Malawi's *Treatment of AIDS* CPG, published in 2008. I began by identifying candidate recommendations from the CPG. I then identified each candidate recommendation's components and evaluated them using criteria for decidability, executability, measurability, computability, and auditability. Each stage of the process identified a subset of the preceding stage to arrive at a final set of recommendations that were suitable for use as performance indicators, given available EMR data from Malawi (Figure 10).

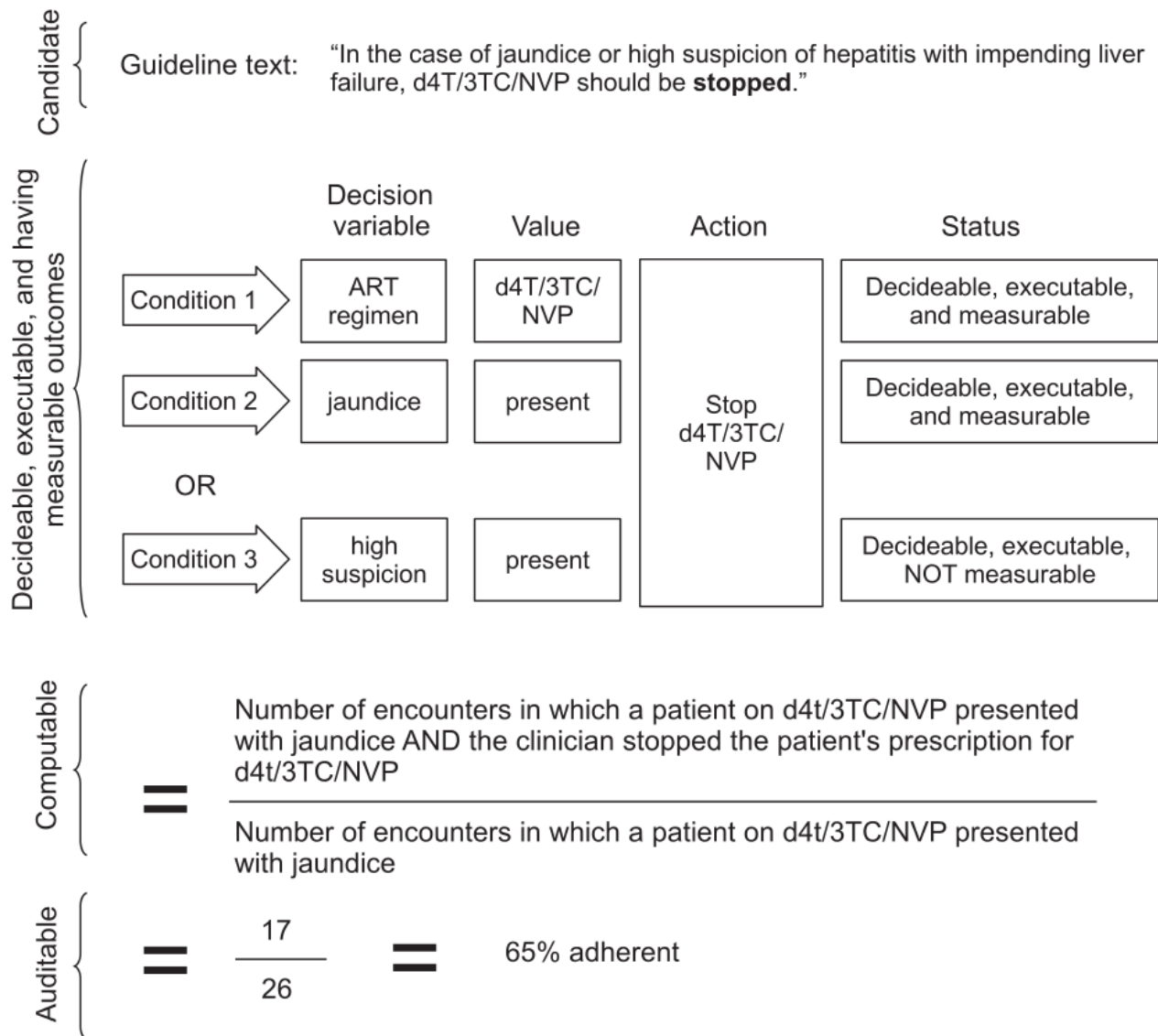


Figure 10: Methods for identification of auditable guideline recommendations.

Candidate recommendations are any statements from the guideline that describe one or more recommended actions to be taken under conditional clinical circumstances. To represent medical knowledge contained in candidate recommendations I used the Guideline Elements Model (GEM), a document-based knowledge representation model for CPGs.¹⁸⁶ I selected GEM for its ability to preserve direct linkages between the CPG text from *Treatment of AIDS*, which the Malawi Ministry of Health mandates all healthcare providers to follow, and the resulting feedback. I used GEM Cutter II, an XML editor that facilitates the mark-up of guideline text to structure recommendations and their components within the GEM hierarchy. For each recommendation, I identified the following components: decision variable, value of decision variable, and action (Figure 10).

To evaluate candidate recommendations for decidability, executability, and measurability, I used the following subset of the criteria from the GuideLine Implementability Appraisal⁷⁶ (GLIA) v 1.0 instrument:

Decidability (precisely under what circumstances to do some thing)

10. If there are more than one condition in the recommendation, is the logical relationship among all conditions (ANDs and ORs) clear?

Executability (exactly what to do under the circumstances defined)

11. Is the recommended action (what to do) stated specifically and unambiguously?

Measurable outcomes (the degree to which the guideline identifies markers or endpoints to track the effects of implementation of this recommendation)

17. Can criteria be extracted from the guideline that will permit measurement of adherence to this recommendation?

Computability (the ease with which a recommendation can be operationalized in an electronic information system)

28. Are all patient data needed for this recommendation available electronically in the system in which it is to be implemented?
29. Is each condition of the recommendation defined at a level of specificity suitable for electronic implementation?
30. Is each recommended action defined at a level of specificity suitable for electronic implementation?

31. Is it clear by what means a recommended action can be executed in an electronic setting, for example, creating a prescription, medical order, or referral, creating an electronic mail notification, or displaying a dialog box?

Two judges, one physician and one clinical officer, independently judged the candidate recommendations using GLIA criteria by scoring them as “Yes,” “No,” or “Unknown.” Where the judges gave discordant scores, they discussed each recommendation to arrive at a final agreed score. Recommendations having any criteria answered as “No” or “Unknown” by both judges were designated as not being decidable, executable, or measurable, and the corresponding barrier(s) to satisfying the criteria were documented. Recommendations having all criteria answered as “Yes” by both judges were assessed for computability.

To assess the computability of decidable, executable, and measurable recommendations, I used the above GLIA criteria for computability.⁷⁶ As a former EMR developer in Malawi, I scored each recommendation as “Yes,” “No,” or “Unknown.” I resolved recommendations that had any criteria answered as “Unknown” by reviewing the recommendation components with another EMR developer. I designated all recommendations that had all criteria answered as “Yes” as computable. Recommendations that did not meet computability criteria were designated as uncomputable, and the corresponding barrier to satisfying the criteria was documented. In answering the computability criterion #28, “Are all patient data needed for this recommendation available electronically in the system in which it is to be implemented?” I documented the required data elements from the EMR that represented each recommendation’s components. I interpreted GLIA criterion #28 to include only data elements that the EMR is potentially capable of collecting, and distinguish these from the smaller set of data elements that are collected in practice. I created a criterion for “auditability” that specifically measures the set of data elements used in practice in the clinics.

Auditability is defined as the availability of representative EMR data for each component of a recommendation. To assess the auditability of each computable recommendation, I created a ratio-based performance measure using the recommendation’s components. The performance measure’s denominator is the number of clinical encounters where specific clinical circumstances were met (the condition). The performance measure’s numerator is the number of clinical encounters where specific clinical circumstances were met (the condition)

and the recommended action was taken.

I selected the clinical encounter instead of the HCW's patients as the unit of measure because of a lack of continuity of care that permits a patient to present the same problems to multiple HCWs over a series of encounters. I created an SQL query representing the denominator of the performance measure for all HCWs in the EMR data. I designated recommendations whose associated denominator query returned zero encounters as being not auditable. For each associated denominator query that returned one or more encounters, I queried the EMR data representing the numerator for an individual HCW to create a proof-of-concept feedback report for the recommendation. I designated recommendations having an associated numerator query returning zero encounters as being not auditable at the cost of excluding any recommendations that were in fact auditable but were universally not adhered to by HCWs during the data collection period. I designated recommendations that had one or more encounters representing the denominator and one or more encounters representing the numerator of the performance measure as being auditable.

For each auditable recommendation, I measured the distribution of average monthly encounters for individual HCWs. I measured average monthly encounters over the entire period of available EMR data for all HCWs who used the EMR for more than one month. I calculated frequency of encounters at a monthly interval based on Jamtvedt et al.'s designation of a monthly interval as a moderate frequency.¹²¹ I analyzed the frequency of recommendations to identify recommendations for which I could more adequately assess HCW adherence, as more frequently occurring recommendations can provide a more reliable indication of HCWs' adherence. However, frequency is not necessarily the only determinant of significance of a recommendation because some rare events may be highly significant.

11.2 RESULTS

I analyzed de-identified, structured EMR data from four ART clinics in Malawi over a 2 1/2-year period from April, 2008 to October, 2010, when the ART guidelines were in effect. During the 2 1/2-year period a total of 117 HCWs (62 clinical officers, 55 nurses) used the

EMR at the point of care to create patient records for 27 528 individual patients. The HCWs recorded a combined 423 831 encounters.

I identified 174 candidate recommendations from *Treatment of AIDS*. Of these, 152 recommendations met the GLIA criteria for decidability, executability, and measurable outcomes. Candidate recommendations that failed to meet the criteria for decidability often related to possible differential diagnoses. An example is the recommendation “Pancreatitis should be suspected if the patient develops severe upper abdominal pain, nausea and vomiting.” This statement failed the criteria for having a measurable outcome because of the difficulty of measuring HCW suspicion. Out of the 152 recommendations that were decidable and executable with measurable outcomes, 58 recommendations met the GLIA criteria for computability. From the 58 computable recommendations, 22 did not have associated EMR data resulting from HCWs’ routine use of the system. Another 15 of the computable recommendations were not assessed for auditability due to complexity in representing temporal constraints. Of the 58 computable recommendations, 21 met the criteria for auditability (Figure 11; totals for the recommendations not meeting GLIA criteria are not mutually exclusive and do not sum up to the n displayed in each box). Proof-of-concept feedback for the 21 auditable recommendations, using EMR data from the month of April, 2010, for one nurse and one clinical officer are displayed in (Table 1).

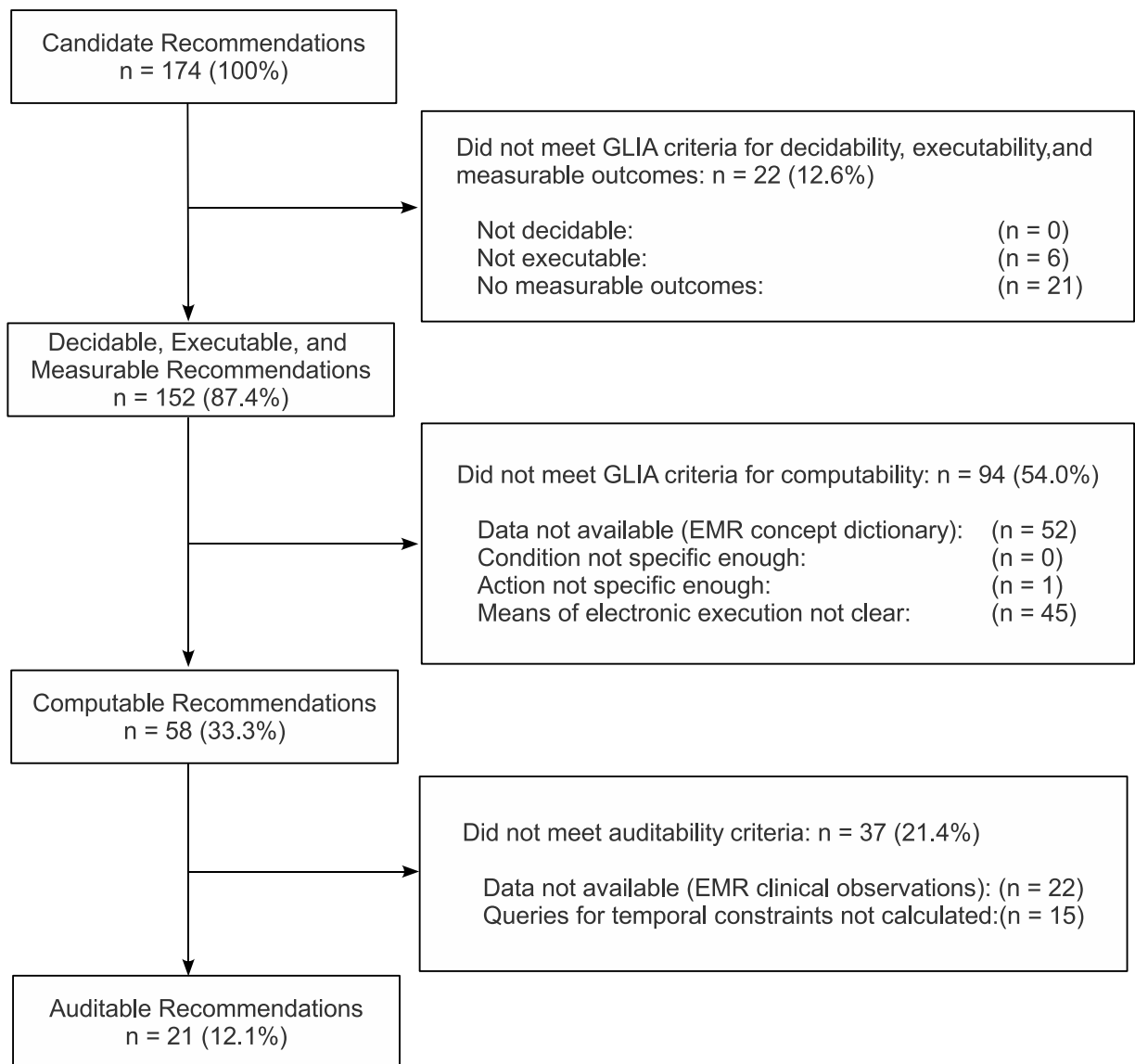


Figure 11: Identification and assessment of recommendations.

Table 1: Auditable recommendations and proof-of-concept adherence feedback.

Auditable recommendations	Sample adherence for April, 2010			
	Individual		Clinic	
	%	Ratio	%	Ratio
Nurse				
If any of the symptoms are recorded as YES, then refer to a clinician.	33.3%	(28/84)	40.4%	(415/1027)
If fever is recorded as YES, then refer to a clinician.	–	(0/0)	62.5%	(40/64)
If abdominal pain is recorded as YES, then refer to a clinician.	75.0%	(3/4)	46.2%	(43/93)
If vomiting is recorded as YES, then refer to a clinician.	–	(0/0)	83.3%	(15/18)
If diarrhoea is recorded as YES, then refer to a clinician.	33.3%	(1/3)	32.4%	(12/37)
If weight loss is recorded as YES, then refer to a clinician.	–	(0/0)	85.7%	(6/7)
If rash is recorded as YES, then refer to a clinician.	11.1%	(1/9)	60.5%	(49/81)
If leg pain or numbness is ... YES, then refer to a clinician.	25.0%	(8/32)	37.2%	(105/282)
If cough is recorded as YES, then refer to a clinician.	21.1%	(4/19)	29.1%	(60/206)
If yellow eyes is recorded as YES, then refer to a clinician.	–	(0/0)	–	(0/0)
If ... changes in body shape is ... YES, then refer to a clinician.	–	(0/0)	100%	(10/10)
If any other new symptom is ... YES, then refer to a clinician.	58.8%	(20/34)	46.5%	(276/593)
If all symptoms are ... NO then the patient can be dispensed ARVs.	92.2%	(329/357)	84.7%	(2757/3256)
Clinical officer				
CPT-eligible patients on CPT	77.3%	(92/119)	89.6%	(1206/1346)
Adult CPT Prescription: One tablet (480mg) twice a day.	100%	(92/92)	100%	(1206/1206)
HIV-positive TB patients will start on cotrimoxazole ...	16.7%	(1/6)	17.8%	(8/45)
Stavudine should not be combined with zidovudine (AZT) ...	100%	(85/85)	100%	(1114/1114)
All adults will now receive the stavudine-30mg regimen ...	100%	(85/85)	100%	(1114/1114)
The drug (Lamivudine) should never be given as monotherapy ...	100%	(92/92)	100%	(1236/1236)
Patients with acute hepatitis (manifested by jaundice) should not be given d4T/3TC/NVP.	–	(0/0)	–	(0/0)
In the case of jaundice ... d4T/3TC/NVP should be stopped.	–	(0/0)	–	(0/0)

ART, antiretroviral therapy; ARVs, antiretroviral drugs; CPT, cotrimoxazole preventive therapy.

The mean frequency for all nurses' auditable recommendations was 45.13, ranging from 0.03 to 580.44 encounters per month. The mean frequency for all clinical officers' auditable recommendations was 58.83, ranging from 0.37 to 341.75 encounters per month. Figure 12 shows the distribution of average encounter frequencies across auditable recommendations. Thirteen of the 21 recommendations (#1-3, 7-9, 12-15, and 17-19) have a mean frequency above four encounters per month, per HCW, making them suitable for providing individualized feedback for at least one encounter per week on average.

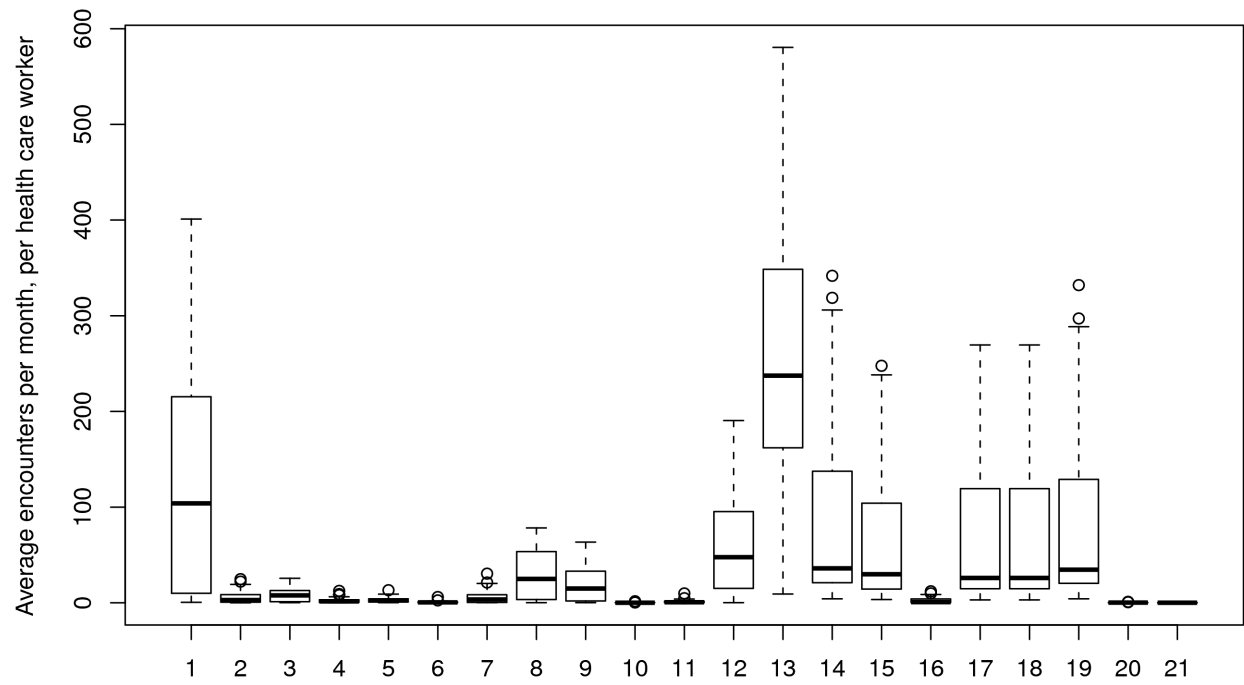


Figure 12: Distribution of average monthly encounters for recommendations in Table 1.

11.3 DISCUSSION

The results of this feasibility study show that 21 (12%) of the recommendations in Malawi's ART guidelines can be audited using EMR data to generate automated performance feedback for an average of 45 clinical encounters per month, per individual HCW. An additional 37 (21%) of the ART recommendations could support AF without requiring alteration of the guideline text or EMR system design. To enable feedback in these cases, 15 out of 37 computable recommendations must be made auditable using representations of temporal constraints to measure HCW performance. The use of temporal reasoning algorithms would allow an automated feedback system to measure adherence to recommendations that are contingent on one or more prior treatment periods, or actions that must be executed within a given time period. For example, a recommendation may indicate that a higher-dose prescription should be given following a month of treatment of the same drug with a lower-dose prescription. The remaining 22 of 37 computable recommendations could not be audited because the data required to represent either the denominator or the numerator were not captured as a part of routine system use. For example, some non-ART prescriptions are routinely recorded on paper, excluding them from electronic audit.

Although 21 recommendations represent a small percentage of Malawi's ART guidelines, performance summaries from the set of recommendations could potentially have a large effect on HCW performance by increasing HCWs' opportunities to reflect on their individual and group performance. Performance summaries of the nurses' referral checklist could provide nurses with new insight into their individual referral and treatment patterns of patients. This feedback can reveal gaps in nurses' understanding of the guideline or differences in beliefs about the efficacy of referral for each symptom. One important symptom is unexplained weight loss, which is associated with early mortality in ART treatment.¹⁸⁷ Another example is patients' complaints of leg pain or numbness indicating peripheral neuropathy, which is highly prevalent but under-diagnosed in Malawi.¹⁸⁸ Clinical officers' performance summaries could similarly reveal prescribing patterns for cotrimoxazole preventive therapy, a priority for AIDS care that can significantly reduce mortality.¹⁸⁹

For ART clinic supervisors, performance summaries can potentially be used to enhance

supportive supervision for HCWs. Clinic-level performance summaries have the potential benefit of facilitating group discussion about HCWs' knowledge of and attitudes toward the guideline. For individual HCWs, performance summaries may assist supervisors in targeting gaps in an HCW's understanding of the guideline. The availability of routine performance data can potentially assist ART clinic supervisors in identifying changing practice patterns over time.

Guideline developers in Malawi may benefit from automated AF data as a form of feedback about the utility of the guideline itself. Using aggregate performance summaries, guideline authors may be able to identify adherence barriers where adherence is uniformly low, or where contrasting levels of adherence occur between HCW groups across ART clinics. I anticipate that the implementation of an automated AF system will contribute value to the process of delivering ART that may in turn increase the demand for feedback data. Providing peer comparison feedback can itself potentially serve as a powerful motivator for HCWs to increase their use of the EMR, enabling a higher percentage of auditable recommendations. Thus there are potential incentives for both EMR developers and guideline authors to develop guidelines and EMR systems to support the delivery of automated AF.

A viable means for increasing the number of auditable recommendations would be a collaborative process for guideline authors and EMR developers to identify the most significant ART recommendations and the EMR data elements that can be collected to represent them using the methods outlined in this study. Using such a process, coordinated development of future versions of the guideline could align ART guideline implementation with automated AF that targets HCWs' lack of familiarity with guideline revisions to accelerate the uptake of new recommendations.

11.4 LIMITATIONS

I excluded recommendations with zero encounters in the performance measure's numerator from the set of auditable recommendations, (e.g. recommendations that no health worker was found to have adhered to over the entire two and a half-year period of data analyzed) because

of the difficulty in distinguishing between true non-adherence and non-use of the EMR by HCWs. Therefore I may have underestimated the number of auditable recommendations. Another limitation of our approach is that the feasibility of providing automated performance summaries to HCWs is dependent upon social, cultural, and environmental constraints that were not evaluated in this study. Future work will evaluate the acceptability of routine performance feedback by HCWs in Malawi to better understand the barriers to providing automated AF in this setting.

11.5 CONCLUSION

A moderate number of recommendations from Malawi’s ART guidelines can be used to generate automated guideline adherence feedback using existing EMR data. Coordinated development of guidelines and EMR systems in Malawi has the potential to increase the feedback that could be generated. Further study is needed to determine the receptivity of HCWs to peer comparison feedback and other barriers to the implementation of automated AF in low-resource settings.

This work establishes the feasibility of automated performance measurement in a low-resource setting using EMR data and a guideline document. This is an essential step for the development of an automated feedback tailoring system because it indicates that there are opportunities to measure performance using automated approaches at the individual provider level. The ability to measure individual-level performance creates the novel possibility of automated feedback tailoring, where this analysis was previously too costly to perform using paper-based records.

12.0 BARRIERS TO USING E-HEALTH DATA FOR PERFORMANCE FEEDBACK IN LOW-RESOURCE SETTINGS

Having established the feasibility of measuring performance in ART clinics using EMR data, my next objective was to identify contextual barriers to the successful use of EMR data for AF in ART clinics in Malawi. I aimed to use the identified barriers to consider how to design technology that supports the provision of individualized performance feedback in hospitals in Malawi.

12.1 METHODS

12.1.1 Setting

The setting for this research is is described in Chapter 3.

12.1.2 Data collection

All data were collected by myself and a junior social scientist at Baobab Health Trust, Mr. Ronald Manjomo. We collected data and performed all other research methods under the guidance of two senior social scientists, one in Malawi and one in Pittsburgh, PA. Data collection occurred between June, 2012 and February, 2013. We conducted activities at eight ART clinics in Malawi's Central and Southern regions. Six of the ART clinics were located in district hospitals and two clinics were in central hospitals. We used the following qualitative methods to collect data about performance measurement and feedback in ART clinics: open-ended interviews, observations, and informant feedback meetings. Participants

were recruited using flyers distributed at each clinic. To protect the rights of participants, we kept research data confidential and did not document identifiable information. The study protocol was reviewed and approved by Malawi’s National Health Sciences Research Committee (NHSRC), in Lilongwe, Malawi, protocol #1019 and by the Institutional Review Board at the University of Pittsburgh, (Pittsburgh, USA), protocol PRO12100159.

12.1.2.1 Interviews: We developed an open-ended interview guide based on our understanding of the clinical setting and on theoretical constructs from two conceptual models (Appendix). We used constructs from the Consolidated Framework for Implementation Research (CFIR) to inform questions regarding feedback in the context of implementation of evidence in clinical settings⁴⁷, and constructs from a cognitive processing model of performance feedback in organizations to inform questions about feedback message processing.^{140,147} The primary feedback-related theoretical constructs we focused on were feedback quality, credibility of feedback sources, acceptance of feedback, perceived accuracy of feedback, desire and intent to respond to feedback, and external constraints that prevented behavior change in response to feedback. We tested the interview guide using two preliminary interviews and revised the interview guide for clarity of language and cultural appropriateness. All interviews were designed to last approximately 30 minutes and were developed to be feasible to complete during clinic time. All interviews were conducted in English and all were audio recorded, except for one interview. For this single interview, data was captured in written notes, then typed into an electronic text file for analysis with other interview data. I transcribed all audio recordings verbatim. We interviewed 32 ART providers, six of whom were clinic supervisors (Table 2). Interview duration ranged from 11 to 39 minutes, with an average duration of 20 minutes. All but one participant agreed to have the interview audio recorded.

Table 2: Characteristics of interview participants (N = 32)

Characteristics	N	%
Sex		
Female	20	63
Male	12	38
Organization		
Ministry of Health	23	72
NGO (Dignitas International)	9	28
Professional role		
Nurse	12	38
Clinical officer	11	34
ART Coordinator	3	9
Nurse supervisor	3	9
Certified nurse-midwife	3	9
Hospital		
District	19	59
Central	13	41
Region		
Central	19	59
Southern	13	41

12.1.2.2 Observations: In addition to the above interviews, we conducted 1-hour observations of healthcare providers using the EMR. We observed seven healthcare providers using the EMR for approximately one hour each. We observed providers in the clinic, before or after holding interviews as time allowed for participants who gave consent to be observed. During the observation, we took field notes about the workflow, system workarounds, and noted where EMR use behaviors may have been associated with data quality problems. We used observations to follow up on what we heard in interviews, and reviewed field notes to inform subsequent interviews and the interview guide.

12.1.2.3 Informant feedback meetings: We met with healthcare providers and supervisors to review preliminary findings at multiple stages during data analysis, with the goal of collecting informant feedback that allowed us to refine our interpretation of the interview data. Informant meetings were held in three district hospitals and at one central hospital. We held three meetings initially after approximately 40% of the interviews had been analyzed, and one meeting after 70% of the interviews had been analyzed. Meetings lasted approximately 30 minutes, with attendance ranging from four to seven participants per site. Additionally, we routinely met and followed up with healthcare provider informants and with representatives from the Department of HIV and AIDS in the Ministry of Health to discuss our findings. Informant feedback meetings were held between May 2012 and July 2013. In meetings with informants we collected field notes that we later used to refine our interpretation of the interview data. We also relied on informant feedback to interpret changes we observed in the EMR software and clinical guidelines over time.

12.1.3 Data analysis:

All interview transcripts were imported into NVivo10 (QSR International Pty Ltd, Doncaster, Victoria, Australia); ZL and RM analyzed the interview data. I constructed a codebook using the editing method described by Crabtree and Miller.¹⁹⁰ All codes emerged from the data in an open, iterative process. I also looked to constructs emerging from our conceptual model and the way they were reinforced by our emerging codes. I maintained an audit trail to

document the creation and refinement of all codes. In addition, each code was clearly defined and provided inclusion/exclusion criteria that helped differentiate it from other codes. As part of the coding process, the first and second author independently coded interviews using the codebook and then met to process any differences in the assessment of codes for each case until agreement was achieved. The codes determined through this adjudication process were then recorded in a master file, which was used for the final analysis.

12.2 QUALITATIVE FINDINGS

Based on our analysis of the interview data, which was informed by field notes from observations and informant feedback, we identified four key barriers to implementing eHealth-based audit and feedback for individual healthcare providers in ART clinics in Malawi: provider rotations, disruptions to care processes, user acceptance of eHealth, and performance indicator lifespan.

12.2.1 Provider rotations

Provider rotations refer to clinic staff schedules that determine how long a healthcare provider works in a clinic. Provider rotations reduce the effect of individualized feedback on performance when a healthcare provider does not stay long enough in a clinic to receive feedback. District hospital clinic participants reported the use of scheduled staff rotations that varied in length from three months to one year.

District hospital providers frequently mentioned rotation schedules:

“I think because we just come here for a few months, ... then you can’t have much experience.” (District hospital nurse-midwife)

“We have adjusted the rotation because in the past we used to have just a week, the other team ... now we have said that each individual should be in the ART at least for three months.” (District hospital clinical officer)

A secondary issue related to provider rotations is an indirect influence on EMR user

training. When user training is not provided frequently enough or in coordination with staff rotations, providers who did not receive EMR user training could potentially create lower-quality data. Several providers raised the issue of not receiving training for the use of the EMR:

“When you come again ... you will not find me here. I will be in another place ... If the training is only done, not quite often, that’s the problem.” (District hospital nurse supervisor)

Although staff rotations were common in district hospital settings, at least one district hospital supervisor reported having permanent staff who did not rotate away from the ART clinic. In contrast to district hospital clinics, central hospital providers did not mention staff rotations, and providers were more likely to work full time in the clinic.

12.2.2 Disruptions to care processes

Disruptions to care processes are unexpected events such as basic infrastructure failures, shortages of pharmaceutical resources, and EMR outages that interrupt or temporarily alter delivery of care. Disruptions represent *external constraints* on performance that reduce the potential impact of feedback, and may lead individuals to perceive feedback as unhelpful for improving performance. Therefore, disruptions that impact clinical behaviors targeted by feedback can represent a barrier to provision of effective individualized feedback. In situations where EMR data are used to monitor performance and the eHealth system itself is believed to be a source of disruption to clinical care, it follows that the *source credibility* and *perceived feedback* of such EMR-based performance feedback is likely to be worse than for disruptions that are not EMR-related.

Disruptions to care processes were common in ART clinics, according to participants. Resource shortages were mentioned by participants as a disruption that interrupted the delivery of recommended care:

“Like at the moment we are supposed to be giving cotrimoxazole, but we don’t have those.”
(Central hospital nurse)

Participants also described disruptions originating from the EMR in the form of system outages. A central hospital clinic had recently experienced system outages that one provider

perceived to prevent the clinic from receiving a quarterly performance award:

“For us, we have been having excellence, excellence, apart from this quarter, where, yeah because of this system, it used to break down, break down, break down, so ... they have seen that we haven’t done well. They haven’t given us the certificate of excellence.” (Central hospital nurse)

Providers also characterized EMR-associated disruptions as minimal in other clinics:

“I just feel that, the system to me, it’s actually good. It does ease the work, yeah. Apart from the disruptions sometimes that are happening, but they are not so common, but with a very high workload, it makes our work actually a bit easier.” (District hospital ART coordinator)

Another type of disruption that we observed when visiting clinics was broken clinic scales that prevented healthcare providers from accurately recording patients’ weight. We received a range of comments about disruptions to care, suggesting that the nature of disruptions may vary according to the hospital setting - district or clinic sites - but that disruptions to care were nevertheless common throughout all of the clinics.

12.2.3 User acceptance of eHealth

User acceptance concerns an individual’s attitudes and intentions towards a technology and his or her actual use of a technology.¹⁹¹ User acceptance is a barrier to the use of eHealth data for performance feedback in that an individual’s complete or partial rejection of the EMR can lead to reduced quality of data for measuring the individual’s performance. This could occur in ART clinics where clinical officers used paper records that were entered retrospectively, instead of using the EMR at the point of care. Participants reported a range of attitudes towards the EMR, and variable system usage patterns. The majority of participants described the EMR as useful and easy to use:

“Using the computer machines has made it simple. I can review so many patients in a minute, unlike using the manual [paper-based system].” (Central hospital nurse)

“The system is working quite OK, and it’s doing a great job to us, looking at the number of patients we are having. It’s easier for us to do the job, rather than to document it ... in

the files.” (Central hospital clinical officer)

In one district hospital clinic, a supervisor indicated that user acceptance among clinical officers was low:

“To be honest, most [clinical officers] are not using it much, most of the data is entered by the clerks and the nurses.” (District hospital ART coordinator)

In this clinic, the clinical officers were reported to use paper records that the nurses entered into the EMR to enable the quarterly clinic-level reports to be generated for national quarterly supervision. As a result, the quality of the data entered was suitable for clinic-level reporting, but would not be adequate for individual-level performance measurement.

Our observations of EMR use also revealed that variable user acceptance of eHealth led to constraints on the ability to use EMR data for performance feedback. ART providers appeared to avoid using some EMR functionality when the EMR workflow did not support established clinical processes, often related to optimizing provision of care under a heavy workload. For example, referral workflow within the EMR was bypassed routinely in district hospital settings, due to the establishment of a more efficient referral process that had not been accommodated by design of the EMR. In central hospital clinics, the EMR workflow appeared to have been configured to match the optimal workflows more closely.

12.2.4 Performance indicator lifespan

The lifespan of performance indicators refers to the average length of time that a performance indicator, once created, remains useful for measuring individualized clinical performance. For example, performance indicators that are based on clinical guideline recommendations may become outdated when a new version of the guideline is published. Malawi published revisions of its national ART guidelines in 2005, 2008, and 2011, and has implemented other planned transitions in recommended practice between publication of a revised guideline. During these transitions some guideline recommendations remained constant, such as those recommending the routine collection of patients’ height and weight. Other recommendations were shorter-lived, such as those that indicated the first-line therapy for new patients, which changed as new, more effective and more affordable drug regimens became available.

The lifespan of a performance indicator, when short (meaning that it is at risk of becoming obsolete within a relatively short period of time), is a barrier to the provision of individualized performance feedback. Because indicators with a shorter lifespan become outdated more quickly, they require more frequent maintenance. Furthermore, using indicators that have shorter lifespans increases the likelihood that feedback based on an outdated measure will be provided, which could reduce the credibility of the technology providing performance feedback.

In Chapter 11 we developed a method for creating performance indicators from statements within a clinical practice guideline document for existing EMR data and identified 21 auditable guideline recommendations that could be used as performance indicators in ART clinics in Malawi.¹⁸⁵ Based on our observations of the clinical setting and follow-up with informants, we observed ongoing changes in two key factors that made these previously developed performance indicators unusable: guideline recommendations and EMR software.

Changes in guideline recommendations occurred as new versions of the guideline were published or new phases of the guideline were implemented nationally in Malawi. We learned that performance indicators that are based on more stable guideline recommendations will require less maintenance and will serve as more reliable indicators of performance over time.

Performance indicators were also impacted by differences in EMR software that we observed in ART clinics in Malawi resulting from ongoing development and implementation of new software versions. As the software is iteratively developed, the addition or removal of functionality can impact the usefulness of a performance indicator. For example, functionality that is used to support the recording of a patient's vital signs is likely to remain unchanged over time, whereas functionality that supports the ascertainment of other clinical signs and symptoms has continued to change to suit the needs of healthcare providers as priorities for screening and referral change within the clinic. In some ART clinics, new versions of the EMR software had not yet been implemented, resulting in differences in the data collected across clinics. We observed that several of the indicators that had been useful for referral and the prescribing of specific drug regimens in Chapter 11 became obsolete due to software changes.

12.3 DISCUSSION

Our goal was to understand barriers to using eHealth data to provide individualized performance feedback in low-resource settings. We identified provider rotations, disruptions to care processes, user acceptance of eHealth and performance indicator lifespan as factors in hospitals in Malawi that could prevent us from generating EMR-based performance feedback. The variability of these factors across hospitals within the same national public healthcare system hold important implications for the design of technology to support the creation and delivery of individualized performance feedback.

As a barrier to providing routine performance feedback, provider rotations in district hospitals appear similar to the problem of staff turnover caused by other factors such as burnout, which has been identified as a barrier to improving the quality of care in other low-income countries.⁹² The provider rotations we encountered in district hospitals in Malawi have been referred to as “cross-training,” which is a capacity building effort that is used to increase provider skills in the management of co-infection for diseases like HIV/AIDS and TB using regular rotations through multiple clinical departments.¹⁹² Considering that the length of staff rotations appeared to be quarterly in several district hospitals, we anticipate that in this setting, feedback that is provided on a monthly frequency would give staff enough time to receive and respond to feedback in their clinical practice before moving on to another department.

Provider rotations in district hospitals also represent a barrier to using eHealth data because shorter rotations can undermine EMR training activities, which in turn can compromise the quality of EMR data created by providers. Resources for EMR training are likely to be extremely limited; therefore, to address the challenge of provider rotations, EMR training should be informed by staff rotation schedules at each hospital. Training interventions delivered via the EMR itself, for example, using interactive tutorials, may successfully address limited training resources in the face of quarterly provider rotations in district hospital settings.

Another implication that provider rotations hold for the provision of individualized feedback is that feedback messages that are provided only within the ART clinic may not be

received by individuals who have rotated to another location. The EMR could potentially be used to address the problem by making feedback reports available in other modules of the EMR, so that providers who have rotated out of a clinic can access routine performance feedback elsewhere.

Our analysis revealed that disruptions to care processes, in the form of resource shortages and EMR technical problems, can reduce the effectiveness of feedback. To accommodate disruptions to care processes in the clinical setting, technology designed to provide performance feedback should include monitoring tools for indicators such as pharmacy stock levels, server uptime, or system usage patterns that could signal when a disruption is likely to compromise the relevance or accuracy of feedback. Disease surveillance approaches may be feasible for use in monitoring eHealth-based disruptions to care processes.¹⁹³ A more practical solution to accommodate the uncertainty created by disruptions to care may be to generate a menu of alternate performance feedback summaries that can be selected or prioritized by a supervisor, based on the supervisor's perceived likelihood of the message to lead to performance improvement. For example, if a drug shortage occurred during a reporting period, a supervisor could choose to prioritize group performance reports that are more relevant than individual performance reports for the current reporting period to reflect the circumstances that were beyond the control of the individual.

Performance indicator lifespan should inform the process of indicator selection and development to avoid wasting resources invested in quality improvement. The most stable performance indicators are likely to be those that are widely used for quality improvement for a specific disease or medical domain. For example, several low-income countries have used measures developed by HEALTHQUAL that focus on simple indicators such as the proportion of patients whose weight was recorded within the month of an ART visit.¹⁹⁴ We found that these indicators were supported by recommendations within Malawi's national ART guideline. The HEALTHQUAL indicators commonly used in low-income countries for ART represent reliable measures of performance that are less likely to become obsolete as guideline recommendations and EMR software change.

Data quality monitoring is likely to be essential for all performance indicators that are implemented for individualized performance feedback. Data quality measures can be used

to estimate the proportion of clinical records that contain errors, and the severity of the errors.^{14,31} When data are routinely assessed prior to the creation of performance feedback reports, an AF system could deprioritize or withhold feedback reports containing unacceptable levels of errors. Furthermore, data quality assessments could be used as a form of performance feedback for providers, encouraging standardized use of the EMR and thereby improving data quality. When individual performance differences are associated with poor data quality, training could be targeted to address the specific providers who have not received instruction about using the EMR, which appears to be a function of provider rotations in many district hospital settings. Data quality assessment programs at each clinic could account for ongoing changes and variability in provider rotations, disruptions to care processes, user acceptance of eHealth, and performance indicators.

The combination of these factors and their variable nature requires automated performance measurement tools to be adaptive to environmental change in low-resource settings, and especially to be able to fail gracefully (e.g., shutting down when problems occur rather than risk generating further errors) by monitoring the presence of errors that may compromise the integrity of performance reports. One promising approach to generating individualized performance feedback in this context is the use of Bayesian methods for the measurement of clinical performance. [40] Furthermore, the ability to generate reports that incorporate recipient beliefs and preferences or which could use a probabilistic approach for estimating the likelihood that feedback will lead to performance improvement are avenues of inquiry that show promise in many clinical contexts.

Variability in provider rotations, disruptions to care processes, and user acceptance of eHealth, and performance indicator lifespan create complexity in the task of adapting to changes over time and across clinical settings. This complexity is increased by differences in providers, who possess different capabilities and motivations to follow the guideline and to use eHealth. The existence of these individual and situational differences are consistent with feedback theories like FPM, which maps the various mediators of the effect of feedback of performance that result from such differences. Feedback which is not tailored to accommodate these factors is likely to be less effective, and in some cases may have a negative effect on performance. Feedback message tailoring is therefore a potentially effective approach to

improving the effect of feedback in this setting.

12.4 LIMITATIONS

Due to the nature of the qualitative data we collected, the results, while important in terms of the information we can gather, are not generalizable. However, for the purpose of identifying design implications for technology that conducts AF, we believe that our methods yielded a sufficient understanding of the key challenges that system developers must overcome in this setting. Another limitation in our approach is that we interviewed ART providers from district hospitals that were located only in Malawi’s Central Region, and from central hospitals located only in the Malawi’s Southern Region. Therefore, any differences between district and central hospitals may be regionally biased. Nevertheless, we believe it is more likely that the differences noted reflect the resource and contextual differences associated with each type of hospital, rather than regional differences. Finally, we collected data only in public hospital facilities in Malawi; therefore, the findings may be less relevant to private hospital settings and clinics with different systems of care in Malawi.

12.5 CONCLUSION

Constant change in clinical real-world systems creates complexity in analyzing eHealth data for performance measurement. To successfully use eHealth data for individual performance measurement and feedback in low-resource settings, technology must accommodate variation in provider rotations, disruptions to care processes, user acceptance of eHealth, and performance indicator lifespan. Technology that enables supervisors to tailor feedback messages for individual and situational differences may improve the effectiveness of performance feedback in this setting.

13.0 DESIGNING KNOWLEDGE-BASED FEEDBACK TAILORING

Having identified complex and dynamic barriers to using EMR data for performance measurement in the clinical environment, I understood that the design of an automated performance feedback system would require a mechanism for accommodating barriers to data use, in addition to accommodating barriers to feedback effectiveness. Based on our qualitative findings, AF evidence, and my understanding of the contribution of psychological theory to the design of behavior change interventions, I explored the feasibility of using psychological theory and knowledge representation methods to design a prototype automated feedback message tailoring system.

Three reasons motivate the use of psychological theory in the design of this system. Firstly, the fact that a large body of evidence (140 randomized controlled trials discussed in Chapter 6) has yet to uncover the mechanisms through which AF influences clinical performance. Secondly, a century of psychological research about behavior and feedback has yielded many candidate causal mechanisms (discussed in Chapter 8) that could be evaluated. Finally, tools that are discussed in Chapter 7 have been developed to facilitate the application of psychological theory for behavior change interventions like AF.

The system design I created is novel in that it supports adaptive message tailoring to specific barriers to behavior change for individual healthcare providers. I proposed that, by leveraging available clinical data, theory-informed knowledge about behavior change, and the knowledge of clinical supervisors or peers who deliver feedback messages, an automated feedback message tailoring system could improve feedback message relevance for barriers to behavior change, thereby increasing the effectiveness of AF interventions. To explore the feasibility the proposed design, I evaluated a selection of theoretical constructs and their implications within a prototype automated feedback tailoring system.

The purpose of this research was to design and evaluate a prototype automated AF system that could facilitate behavior change within complex clinical environments. I designed the system based on the following assumptions: First, I assume that performance feedback is being given routinely in clinical settings to healthcare providers for the purpose of knowledge translation, including quality improvement and the implementation of evidence-based practice. Second, I assume that clinical supervisors have some awareness of individual healthcare providers' barriers to change. For example, a supervisor may believe that an individual's low performance is caused by a lack of motivation rather than lack of knowledge or skill. Third, I assume that clinical supervisors heuristically or intuitively tailor feedback messages to some degree, whether verbally or in writing, when giving feedback to healthcare providers. For example, a supervisor may use the "feedback sandwich" technique to deliver a feedback message by "sandwiching" negative feedback between two positive feedback messages.¹⁴² Finally, I assume that the quality of performance data in some cases is adequate to convey meaningful performance feedback to healthcare providers.

It is under these conditions, when supervisors interpret credible performance data based in part on their beliefs about individual healthcare providers' barriers to behavior change, that I envision a feedback message tailoring system to have a significant and positive influence. I anticipate that such a system, if designed appropriately, could help to address pain points for clinical supervisors, healthcare providers, and AF researchers. I envision such a tool as being helpful for clinical supervisors who dislike giving feedback or who would benefit from having a range of theory-informed, recommended feedback messages to choose from for each specific provider. For healthcare providers, I anticipate the system would increase the provision of relevant feedback and decrease the amount of useless or harmful feedback coming from a supervisor or peer. Furthermore, I envision such a system to enable AF researchers to observe the tailoring of feedback messages and use of differing AF components under heterogeneous and dynamic conditions, to generate knowledge about the effectiveness of AF.

The approach I created differs from prior work on AF interventions in several ways. First, studies of AF have used tailoring to adapt an intervention to a local context, for example a country, institution, or a specific clinic. I differentiate the type of tailoring I am describing as

being about the design of specific feedback messages, created by a supervisor or peer using a software application, for each individual rather than for a group of providers. Second, the design of feedback in AF interventions is typically established prior to the intervention and remains constant throughout, but I am proposing a mechanism for the continued adaptation of feedback report design prior to the delivery of each message. Third, studies of AF have explored the optimal design of feedback messages, such as comparing the effect of graphical vs textual information, or delivery of messages in writing vs in person.¹⁰⁷ This important work however has not evaluated messages designed for individual providers rather than for the recipient population. Fourth, many studies of AF provide a static report to providers that includes a consistent set of performance measures, such as multiple process and outcome measures of antibiotic prescribing behavior. While this kind of report is completely relevant to our discussion, I discuss message tailoring at the level of each performance measure, for example prioritizing one measure over another, conditional on factors that are most likely to lead to performance improvement, rather than sending an static set of indicators to all providers. Finally, I bound the scope of our discussion to address routine, unsolicited feedback messages sent to a healthcare provider, excluding feedback provided informally outside of the feedback intervention context (e.g. in response to feedback-seeking behavior¹⁹⁵).

13.1 METHODS

To determine the feasibility of using theoretical constructs to guide feedback message tailoring, I conducted theoretical modeling work in consultation with an AF theorist who is a cognitive psychologist. I used a menu of constructs approach with a range of example constructs that are intended to sufficiently support my argument, but not to definitively survey the theoretical landscape. Using the example domain of antimicrobial stewardship, I mapped theoretical constructs to barriers to antibiotic prescribing to demonstrate the relevance of some constructs to hypothetical causal mechanisms that performance feedback may leverage. My goal in using these examples was to describe a range of constructs and the tailoring actions that could be taken for each construct, showing how feedback might impact

behavior differentially according to specific barriers to behavior change.

Based on the results of this modeling work, I designed a knowledge-based information system and preliminary knowledge base that uses the causal mechanisms and theoretical constructs we identified for antimicrobial stewardship. I formulated a message tailoring process and identified requirements for an information system that could operationalize the message tailoring process.

13.2 RESULTS

I selected six examples of TDF constructs, each from a different domain, with two constructs mapped to each of the three COM-B categories (Table 3). Each construct contains one or more hypothetical causal mechanisms that hold implications for AF tailoring based on individual or situational characteristics. These six constructs and the causal mechanisms they offer are examples from what I anticipate is a broader set of constructs yet to be identified that could be used to guide feedback message tailoring.

Knowledge (including knowledge of condition/scientific rationale) is a construct from the “Knowledge” TDF domain, defined as “An awareness of the existence of something.”¹³⁶ “Knowledge of a condition or scientific rationale for a behavior” as a barrier to behavior change can be directly impacted by a feedback message when the recipient lacks the targeted knowledge. Feedback will be less relevant when provided to an individual who already has the targeted knowledge about performance or the behavior. For example, an intervention to improve unnecessary antibiotic prescribing, deciding not to prescribe antibiotics requires providers to know the specific clinical conditions for delay of prescribing. Providers who already know the conditions will find feedback about performance less relevant with regard to this narrow dimension of the prescribing behavior.

Interpersonal skills is a construct in the “Skills” TDF domain, defined as “An aptitude enabling a person to carry on effective relationships with others, such as an ability to co-operate, to assume appropriate social responsibilities or to exhibit adequate flexibility.”¹³⁶ Interpersonal skills are important as a capability barrier that, if salient, are unlikely to be di-

rectly affected by performance feedback. In the case of antibiotic prescribing, where patient demands are a barrier to behavior change, poor interpersonal skills may cause a provider to acquiesce to a patient's demands for antibiotics, because the provider feels ill-equipped to deny the patient a prescription, at the risk of damaging the patient-provider relationship. In this case, training is more likely to lead to improved provider capability that enables behavior change, whereas repeated negative feedback about poor performance could potentially reinforce a provider's beliefs about lack of interpersonal skills, worsening future performance.

Material resources is a construct in the "Environmental Context and Resources" domain of the TDF, defined as "commodities and human resources used in enacting a behavior."¹³⁶ Material resources are associated with feedback in that recipients who lack resources necessary to enact a behavior are likely to find performance feedback less relevant, whereas recipients with adequate resources are likely to find feedback to be more relevant. Additionally, an intervention that involves resource stewardship could be confounded by resource shortages that artificially improve performance, resulting in the provision of less relevant performance feedback that shows performance improvement. For example, in an antibiotic prescribing intervention that uses a restrictive approach such as implementing an expert approval requirement, performance feedback about the constrained behavior is likely to be less relevant to the individual because performance feedback has less bearing on the clinical decisions made in any specific situation.

Table 3: Mapping antibiotic prescribing barriers to theoretical constructs.

COM-B category ¹³⁸	TDF domain ¹³⁶	TDF construct ¹³⁶	Barrier to antibiotic prescribing ¹⁰⁹	Hypothetical causal mechanism
Capability	Knowledge	Knowledge of condition/scientific rationale	Lack of knowledge and training	Feedback can change awareness to impart new knowledge that leads to behavior change
	Skills	Interpersonal skills	Perception of patient demands and preferences	None (Feedback has no direct influence on interpersonal skills)
Opportunity	Environmental context and resources	Material resources	Inadequate drug supply infrastructure	None (Feedback has no direct influence on material resources)
	Social influences	Social pressure	Peer pressure and social norms	None (Feedback has no direct influence on social pressure)
Motivation	Beliefs about capabilities	Self-efficacy	None (Barriers are indirect via beliefs about capability constructs)	Feedback can influence perceptions of ability, improving or worsening self-efficacy, which can lead to behavior change
	Emotion	Fear	Fear of bad clinical outcomes	Feedback can cause emotional reactions that influence motivation, leading to behavior change

Social pressure is a construct in the “Social influences” domain of the TDF. Social pressure is defined as “the exertion of influence on a person or group by another person or group.”.¹³⁶ The construct of social pressure is important for feedback effectiveness as a situational characteristic that could indicate when peer comparison feedback should be used. As group performance changes from low to high, the presence of social pressure, if salient, could be presumed to influence individuals to move toward the group performance mean. When group performance is low, peer comparison feedback showing a low group mean could potentially reinforce negative social pressure. Therefore to improve the effect of feedback on performance, social pressure could be accommodated by withholding comparative feedback until a significant percentage of the group had achieved a high level of performance.

Self-efficacy is a construct in the “Beliefs about Capabilities” domain of the TDF. Self-efficacy is an individual’s perceived ability to control their own performance and the events that affect them, using the resources they have at hand.^{196,197} In cases where self-efficacy for a given task is low, repeatedly negative feedback or peer comparison feedback showing diminishing performance relative to peers may worsen the recipient’s self-efficacy.¹⁹⁸ This could lead an individual more quickly towards goal abandonment rather than increased effort to improve performance. For example, consider a physician participating in an AF intervention for antibiotic prescribing who believes he has poor patient communication skills, as evidenced by his patient experience survey scores. For this physician, the perception of patient preferences and demands may represent a formidable barrier to improving performance. If the physician does not improve, showing the physician repeatedly negative or declining performance scores for prescribing behaviors could lead the physician to have lower self-efficacy for antibiotic prescribing tasks, motivating avoidance behaviors rather than motivating improved performance. A more appropriate solution under the circumstances could be to emphasize relative improvement rather than comparative performance gaps. Another potential tailoring solution would be to withhold repeated and very negative feedback, and instead offer the low-performing physician a refresher training course, or to seek structural changes that could facilitate performance improvement.

Fear is defined as “An intense emotion aroused by the detection of imminent threat, involving an immediate alarm reaction that mobilizes the organism by triggering a set of

physiological change.”.¹³⁶ Fear is a construct in the “Emotion” domain of the TDF. When the construct of fear is salient, feedback messages could interfere with perceptions of feedback, depending on the emotions of the provider and their perception of the behavior. For example, if a provider is afraid that declining to prescribe antibiotics will lead to a bad outcome, feedback showing poor clinical outcomes could trigger physiological changes that prevent the provider from perceiving other meaningful feedback indicators on a feedback report. In this case, feedback could have high personal relevance for the provider, but not be effective for improving performance.

The group of constructs I discuss above spans all three COM-B categories, and six of the 13 TDF domains. I anticipate that this sample is a small proportion of the set of constructs that could potentially be used to tailor feedback messages. A feedback tailoring system could use the above constructs to guide the automated tailoring of performance data for many possible barriers. To illustrate how these constructs could be operationalized within an automated message tailoring system, I use a scenario in which a clinical supervisor is preparing to give feedback to a low-performing physician below in Section 13.2.1). Based on the performance data and information about the behavior, a message tailoring system could create a range of graphical and textual messages that the supervisor could review and select, according to her perceptions of the specific barriers to behavior change for the physician (Figure 13).

13.2.1 Low-performance of unnecessary antibiotic prescribing for acute respiratory infection

Performance is routinely measured for physicians participating in an antimicrobial stewardship program in a hospital setting. To measure unnecessary prescribing of antibiotics for patients diagnosed with acute respiratory infection (ARI), an inverse proportional measure of prescribing behavior (0% is completely compliant) is calculated as follows:

Numerator Number of patients diagnosed with ARI for whom antibiotics are not indicated
AND antibiotics were prescribed

Denominator Number of patients diagnosed with ARI for whom antibiotics are not indi-

cated

Dr. A is a supervising physician who is responsible for implementing the antimicrobial stewardship program. She uses the above performance measure to calculate individual performance for each provider in her department. She also calculates a combined average for the top 10

Dr. B is a physician who has performed consistently low, relative to his peers, over the previous year. Dr. A is preparing performance feedback to review with Dr. B, and she has calculated the performance data in Table 4.

Table 4: Individual antibiotic prescribing performance data with a peer benchmark

Quarter	Performance	Top 10%
2013 Q3	86.5	47.5
2013 Q4	84.3	46.2
2014 Q1	85.9	41.3
2014 Q2	80.1	38.8

Dr. A must decide how to present the above data to Dr. B, in a way that is most likely to lead to performance improvement. In this scenario, we assume that individualized performance feedback about antibiotic prescribing behavior for ARI can and should be provided for the following reasons:

1. Performance barriers for the behavior are associated with individual physicians (eg capability and motivation) rather than situational constraints, and Dr. A is not aware of any disruptions during the recent reporting period that would have influenced performance for this measure.
2. Antibiotic prescribing behavior for ARI is not a team-associated behavior (eg does not require significant task coordination across providers) therefore individual performance feedback for this behavior is more relevant than group feedback.
3. Dr. A has assessed the quality of the clinical data used to measure performance, and believes that the level of accuracy in performance measurement is acceptable.
4. Behavior change for this measure is evidence-based and achievable for all providers, and is therefore a clinical quality improvement priority.

The feedback messages in Figure 13 are contrasting versions of the same performance data about Dr. B from 4. To use this information, a supervisor or peer of Dr. B could identify a description in the leftmost column that most closely matches their own beliefs to identify a theory-informed, tailored feedback message in the rightmost column that is relevant to Dr. B's specific barriers to behavior change.

Figure 13: A menu of tailored messages for a low-performing healthcare provider.

Supervisor or peer beliefs about recipient	Salient constructs	Tailoring approach	Message	
Recipient is aware of similar past performances. Recipient may lack required interpersonal skills, and could benefit from training. The recipient may have responded negatively to feedback in the past.	Self-efficacy, interpersonal skills, fear	Withhold feedback to avoid worsening self-efficacy or fear; facilitate self-reflection; recommend training	[None]	Formative / Professional development-focused
Recipient may be unaware of current performance.	Knowledge	Show score only, facilitate self-reflection	80.1%	
Recipient is aware of similar past performances. Recipient may believe he is not capable of improving performance. Increased effort is likely to lead to improvement for the recipient.	Self-efficacy	Emphasize recent improvement using self-comparison and a truncated vertical scale		
Recipient may be unaware of his current performance and/or peer performance for the behavior. Recipient has adequate skill required to improve performance. Recipient is influenced by peer performance.	Social pressure, knowledge	Compare current performance with that of top-performing peers to motivate improvement		Summative / Corrective action-focused
Recipient may be unaware of differences between his current and past performance and that of peers. Recipient has adequate skill required to improve performance. Recipient is influenced by peer performance.	Social pressure, knowledge	Compare performance history with that of top peers to motivate improvement		

13.2.2 Automated feedback message tailoring

The examples of tailored feedback in Figure 13 could be developed by a software algorithm that identifies features of an individual provider’s performance and then creates a range of possible messages from which a supervisor could select. To determine how to tailor a message for a recipient, the system could use facts about the performance data, theoretical constructs, and the clinical context. For example, to generate the graph in the middle of Figure 13 featuring the truncated scale, the system would need to “know” the following facts:

1. Low performance for this behavior in this setting is any value less than 70% (context-specific performance threshold)
2. Repeated low performance may be associated with low self-efficacy (performance feature-construct relationship)
3. Self-efficacy is a construct that can be positively influenced by feedback messages that emphasize improvement (causal mechanism, e.g. construct-message feature relationship)
4. Truncated scale graphs can be created when current performance shows an improvement of more than 5% over previous performance (performance feature-tailoring action relationship)
5. Truncated scale graphs potentially can be used to emphasize improvement (Tailoring action-message feature relationship)

Creating a knowledge-base that contains facts like those above is a key step in the development of a system for feedback tailoring. A key challenge for creating a feedback tailoring knowledge base is to develop a valid classification of feedback message elements. Defining and understanding these elements will involve the development of other novel forms of knowledge representation for AF interventions. For example, I do not know the set of message tailoring actions (e.g. graphical scale truncation, withholding, prioritizing, message-based psychological priming) that are meaningful for AF interventions. Much of this work could build directly upon ongoing efforts to formalize terminology for intervention specification and reporting^{199,200}, and frameworks that facilitate the systematic use of theory, like the TDF and COM-B. I view the broader formalization of theory-informed implementation knowledge

as a foundation for the development of computer-interpretable message tailoring knowledge-bases.

13.2.3 Using performance features to indicate construct salience

To tailor performance feedback for a specific causal mechanism, one must estimate when a theoretical construct that contains the causal mechanism is salient. For example, to know if feedback is likely to change awareness by imparting new knowledge that leads to behavior change for an individual provider, one needs to determine that knowledge is a salient construct, meaning that lack of knowledge is an actual barrier to behavior change for the individual. A central proposition I make in demonstrating the feasibility of automated AF tailoring is that features of an individual's performance data (e.g. individual and situational characteristics of a provider, including past performance, peer performance, provider role, etc) can be used to estimate the salience of a theoretical construct as a determinant of clinical behavior.

Performance features are the individual and situational characteristics associated with an individual provider, and his or her behavior that is targeted by an AF intervention. Each performance feature has a relationship with one or more constructs. My objective is to present a range of performance features to sufficiently demonstrate the feasibility of using performance features to indicate construct salience. Based on the causal mechanisms we identified in the theoretical constructs from the TDF, I developed a preliminary feedback tailoring knowledge base contains the following performance features:

1. The provider never performed the task (numerator = 0)
2. The provider has one consecutive prior month of performance data that can be used for comparison
3. The provider has two consecutive prior months of performance data that can be used for comparison
4. The provider treated more than four patients (denominator > 4) during the current month

5. The provider's peers performed above 90% on more than two days that the provider was working during the current month
6. The provider's average performance is 10% below the average of the two top performing peers
7. The provider's average performance is above 90% for current and two prior consecutive months
8. Individual's performance changed by more than 5% compared with last month's performance
9. Provider's average performance is above 90%
10. Provider's average performance is in the top 25% of performances (upper quartile)
11. Average performance for all providers is below 50.

The above set of features consists primarily of information about performance variability and performance history. *Performance variability* refers to features of change or stability in performance of one or more providers. Performance variability of individual providers could be used to model the influence of constructs that are believed to be strongly associated with specific patterns of performance variability. For example, when knowledge is a barrier to antibiotic prescribing, an individual is unlikely to have consistently high performance. Conversely, individuals who have consistently low or inconsistent performance are more likely to lack the knowledge required to prescribe antibiotics appropriately. I propose that by using simple thresholds at reasonable bounds of high, moderate, and low performance to estimate the salience of some constructs, feedback could be tailored to be more effective than non-tailored feedback. I note that the thresholds used to determine high, moderate and low performance are subjective and must be adjusted for expectations and contextual factors. For example, an AF intervention to reduce antibiotic prescribing may involve physician assistants, family physicians, resident physicians, and specialists, all of whom are employed at an academic medical center. Because of differences in patient risk for disease and limitations in the number of clinical variables monitored, performance variability may be expected to be higher for specialist physicians than for physician assistants. Therefore the performance feature thresholds indicating "high" and "low" performance for one provider role may not apply

to all provider roles, and may need to be established based on past performance and the expectations of clinical supervisors with knowledge of the intervention setting.

In addition to performance variability features, an individual's *performance history* could be used to indicate the salience of constructs. For example, if an individual has consistently high past performance, the likelihood that knowledge is salient as a construct is low for the current performance summary, even if the individual's current performance is low. On the other hand, if an individual has never demonstrated high performance but peers have demonstrated high performance concurrently in the individual's performance history, the salience of all capability and motivation constructs is likely to be higher than opportunity constructs for the individual. Performance history features could also be used to determine if an individual has received repeated feedback that is consistent, yet has not responded by changing behavior. Repeated consistent feedback could be potentially used to indicate a lack of salience for associated constructs. For example, if an individual performs consistently low, yet has received and viewed repeated feedback messages indicating low performance and containing information targeting knowledge gaps, this scenario could indicate that lack of knowledge is not a barrier to behavior change.

Beyond features of performance variability and performance history drawn from the performance data alone, alternate existing sources of data might be used to inform message tailoring. Data that holds implications for the salience of constructs may be available in many clinical settings, including pharmaceutical inventory data, patient experience scores, and provider's human resources records. For example, hospitals that use electronic prescribing may have pharmaceutical inventory status reports that could indicate the salience of material resources based on drug shortages. Patient experience scores and other clinical quality data, which are increasingly being made public at the individual provider level²⁰¹, may contain provider characteristics that could indicate the salience of capability constructs such as interpersonal skills. Provider experience could be associated with adequate knowledge and thereby a lack of salience for knowledge as a barrier to appropriate use of antibiotics.²⁰² Provider role information could be used to estimate the salience of peer influence, where team members have differentiated roles (e.g. senior physician, junior physician, physician assistant, infectious disease specialist) based on belief-based models of level of peer influence

across roles.

Providers who are interested in receiving tailored feedback may be willing to provide information about their perceptions and experiences that could directly indicate the salience of theoretical constructs. For example, providers may be willing to complete a questionnaire containing a self-efficacy scale about prescribing behavior.²⁰³ Perceived barriers to behavior change could be reported directly from providers that could indicate the salience of specific constructs. Provider-reported beliefs about antibiotic prescribing could be used to indicate the salience of motivational constructs.^{204,205}

13.2.4 System architecture

I have developed a general system architecture for a knowledge-based feedback tailoring system (Figure 14). Knowledge-based systems encode expert knowledge to support automated reasoning about a set of facts maintained within a knowledge base.¹⁶³ The approach to reasoning that I outline here uses if-then rules for the purpose of a) inferring the salience of theoretical constructs related to the cognitive processing and behavioral responses to performance feedback and b) selecting tailoring actions within a message tailoring process to increase the likelihood of its effectiveness. Our goal in presenting the system architecture is to demonstrate how our conceptual model could be operationalized and thereby to motivate research addressing the potential significance and impact of theory-informed AF tailoring systems in the context of implementing evidence-based practice. The system architecture has five major components: A tailoring knowledge-base, an eHealth database, a performance measurement process, a performance database, and a message tailoring process.

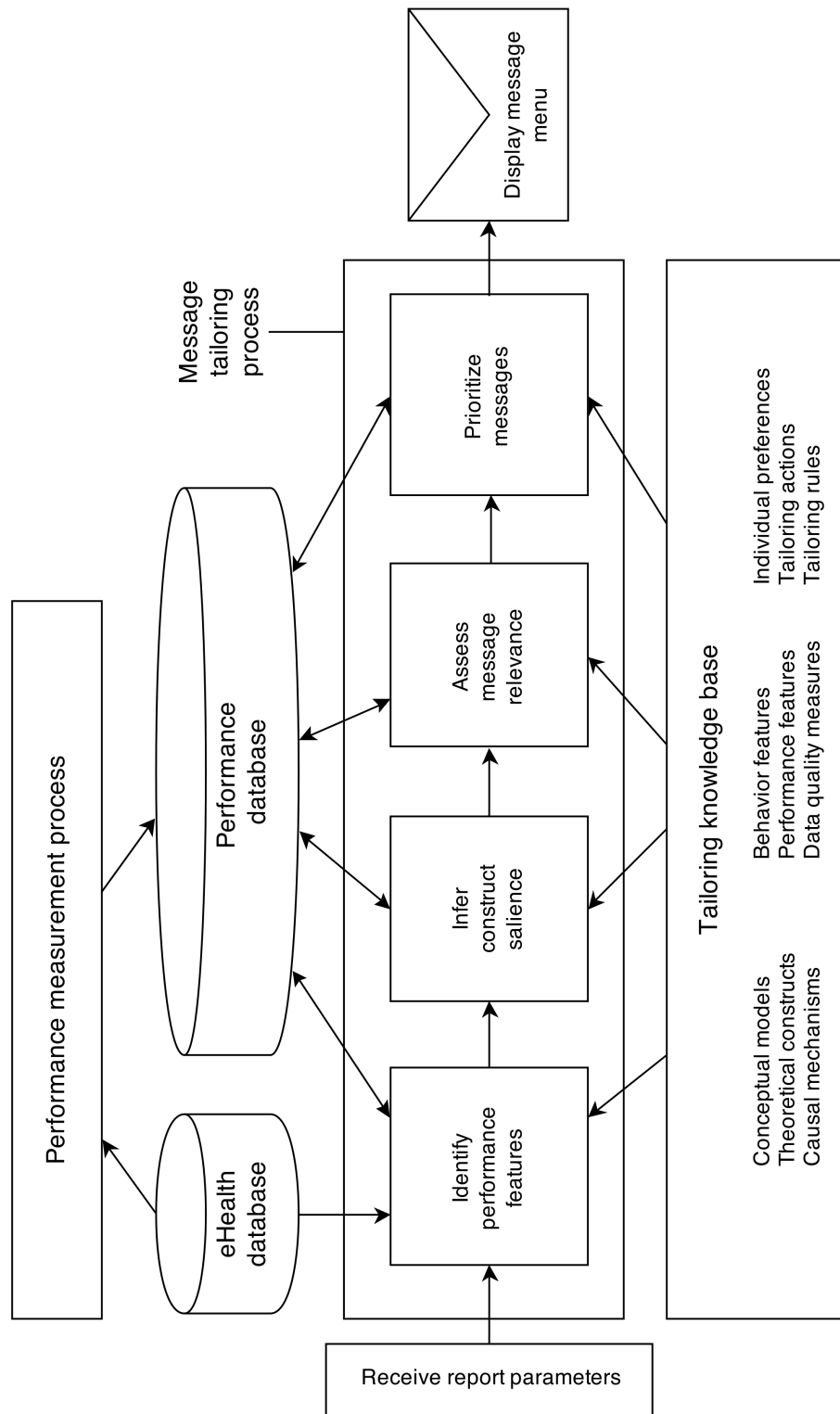


Figure 14: System architecture for an automated AF message tailoring system.

13.2.4.1 Tailoring knowledge base A knowledge base is a store of information that contains a collection of facts and rules that a program can use to perform complex reasoning about a situation or event. The AF tailoring knowledge-base models expert knowledge based on psychological theory and the medical domain of the targeted behavior, and knowledge in the form of local expectations and beliefs of healthcare providers. This local knowledge is required to fit the intervention to the environment in which the system is to tailor performance feedback. As such, the configuration of a tailoring system requires in-depth understanding of the local clinical environment. The knowledge base I describe contains the theoretical models and conceptual frameworks, constructs, performance features, behavior features, and other concepts that could allow such a system to reason about how to provide performance feedback optimally (Table 5).

Table 5: Definitions of tailoring knowledge-base components.

Component	Definition	Examples
Conceptual framework	A model that can be used to organize, manage, select and make inferences about theoretical constructs that hold implications for AF tailoring goals	Cognitive processing model of performance feedback (Ilgen, Fisher, and Taylor 1979); Capability, Opportunity, Motivation, and Behavior (COM-B) (Michie et al 2011); Theoretical Domains Framework (TDF) (Michie et al 2005)
Theoretical construct	An explanatory variable that is a components of broader theories of behavior and cognition that hold implications for AF tailoring	Knowledge (including knowledge of condition/scientific rationale), interpersonal skills, material resources, social pressure, self-efficacy, fear
Causal mechanism	A path of influence between factors within a theoretical construct that has meaningful implications for feedback interventions	Feedback can change awareness to impart new knowledge that leads to behavior change; Feedback can influence perceptions of ability, improving or worsening self-efficacy, which can lead to behavior change
Behavior feature	A characteristic of a targeted behavior and its component behaviors or tasks that can be used to make inferences about the salience of theoretical constructs	Unnecessary prescribing performance can be artificially improve by stockout of antibiotics; Interpersonal skills are required as an indirect determinant of unnecessary prescribing behavior because of the need to convince patients to accept delay of prescribing without damaging the patient-provider relationship
Performance feature	A characteristic of individuals and situations that can be used to infer or estimate the salience of theoretical constructs	Performance variability across providers and provider groups, consistency of past performance, known resource shortages, provider experience, patient experience scores, provider reported information, provider role
Data quality measure	An indicator of the degree to which available eHealth data is fit for the purpose of indicating performance	Number of patients who were prescribed antibiotics is not greater than the number of prescriptions created; Duration of prescription is within an expected range
Individual preference	Reported partiality or affinity towards specific attributes of a feedback message	Process vs outcome measures, peer comparison, delivery channel (email, web, SMS), visualization preferences that accommodate color-blindness
Tailoring action	An activity that involves the transformation of a feedback message to improve its relevance or likelihood of leading to performance improvement	Prioritizing: ordering messages or emphasizing their importance; Framing: presenting or visualizing information to change the emphasis of feedback; Withholding: suppressing or excluding components of a feedback message
Tailoring rule	An if-then statement that evaluates performance and behavior features to select tailoring actions when conditions are satisfied	If past performance is consistently low and current performance is low, withhold peer comparison feedback showing high peer performance to avoid damaging self-efficacy

Behavior features are theory-informed attributes of a behavior that the intervention is targeting, and are also used as components of conditions within tailoring rules. The attributes of the behavior include COM-B components and other features relevant to tailoring in a specific context, such as the relative priority of the behavior and the professional roles of providers who perform the behavior.

Tailoring actions are the set of possible transformations that a tailoring system can apply to performance information. In the theoretical constructs and tailoring examples discussed above, I identified the following tailoring actions:

Framing: Framing refers to the emphasis and tone that a feedback message uses. For example, framing can be used to change the emphasis of a feedback by including or excluding comparison information, changing the tone of the language, or changing the scale of a chart's axis to emphasize change or trends over time.

Priming: Priming can influence the psychological state of the feedback recipient to increase receptivity to a feedback message. An example of a priming technique that is commonly practiced is the “feedback sandwich.”.¹⁴²

Prioritizing: Prioritizing becomes increasingly important as the number of potential feedback messages increases. I anticipate that prioritizing feedback according to data quality and message relevance is an essential tailoring action for improving the effect of feedback messages on behavior.

Withholding: When it can be determined that there is a high likelihood that feedback may cause the opposite reaction to that which is intended, components of feedback messages should be withheld. Feedback that is likely to encourage goal abandonment, (e.g. continued extremely negative feedback) is unlikely to result in improved performance and may even result in negative effects on performance. If it could be determined that a particular situation is leading to continued, extremely negative feedback, withholding the feedback message may be preferential to delivering the message.

Comparison: Performance feedback can be displayed with or without comparison to a goal, standard, peer performance, or expert performance. Theoretical constructs like self-efficacy suggest that performance comparisons can be motivating or demotivating to an individual depending on situational features.

Tailoring rules are if-then statements that evaluate performance and behavior features and used weighted scoring to estimate the salience of constructs or relevance of message components when the conditions of a rule are satisfied. Tailoring rules represent an interpretation of one or more causal mechanisms offered by theoretical constructs. For example, a causal mechanism offered by the construct of self-efficacy claims that repeated negative feedback compared to concurrent high peer performance can worsen self-efficacy, leading to lower motivation. To represent this mechanism, interpretation must be done to fit the mechanism to the intervention context. The definition of negative feedback for this case depends on the expectations of a threshold for low performance for each specific behavior, and perhaps for specific provider roles or patient populations.

I created the following tailoring rules that the system uses to estimate the salience of theoretical constructs and the relevance of message components, based on the presence and absence of performance features:

1. If the provider does not have two prior consecutive month of performance data that can be used for comparison, then set the relevance score for self-comparison and peer-comparison-historical to 50 (less relevant).
2. If the provider never performed the task (numerator = 0), and the provider treated more than four patients (denominator > 4) during the current month, and the provider's peers performed above 90% on more than two days that the provider was working, then increase the estimated salience of capability barrier constructs (e.g. knowledge +=1, skills +=1) and increase the relevance score for peer-comparison component (comparison score += 1).
3. If the provider's average performance is > 10% below the average of the two top performing peers, then increase the estimated salience of capability barrier constructs (e.g. knowledge +=1, skills +=1) and increase the estimated salience of motivation barrier constructs (e.g. self-efficacy +=1), and increase the relevance score for peer-comparison component (comparison score += 1).
4. If the provider has one prior consecutive month of performance data that can be used for comparison, and the individual's average performance changed by more than 5% compared with last month's average performance, then increase the relevance score for

self-comparison component (self-comparison score $+= 1$) and increase the relevance score for scale-truncation component (scale-truncation score $+= 1$).

5. If the individual's average performance is above 90% for the current month, then decrease the estimated salience of capability barrier constructs (e.g. knowledge $-=1$, skills $-=1$), decrease the estimated salience of motivation barrier constructs (e.g. fear $-=1$, self-efficacy $-=1$), decrease the estimated salience of opportunity barrier constructs (e.g. material resources $-=1$, peer pressure $-=1$) and decrease the relevance score for peer-comparison component (comparison score $-=1$)
6. If the individual's average performance is at or above 75th percentile, then decrease the relevance score for peer-comparison component (comparison score $-=1$)
7. If the individual has two prior consecutive month of performance data that can be used for comparison and the individual's average performance is 90% or higher for current and two prior consecutive months, then decrease the estimated salience of capability barrier constructs (knowledge $-=1$, skills $-=1$)
8. If the average performance for all providers is below 50% then decrease the estimated salience score for capability barrier constructs (knowledge $-=1$, skills $-=1$) and increase the estimated salience of opportunity barrier constructs (e.g. material resources $+=1$, peer pressure $+=1$)

13.2.4.2 eHealth database The eHealth database contains patient medical records, clinical information systems data, hospital administrative records, and other sources of data that could be analyzed to determine the present or absence of performance features for each provider's performance.

13.2.4.3 Performance measurement process The performance measurement process uses performance measures based on a clinical guideline to analyze eHealth data and create performance reports in the performance database. For example, the method we developed in Chapter 11 could be used to generate the performance data, or it could be created using a manual chart review that abstracts performance data from clinical records.

13.2.4.4 Performance database The performance database contains each individual's performance data, associated metadata about performance measures, provider profiles, provider performance reports, and the message tailoring assessment data that is used in the message tailoring process.

Table 6: Message tailoring process from a prototype feedback tailoring system.

Steps	1. Identify performance features	2. Infer construct salience	3. Assess message relevance	4. Prioritize messages
Method	Binary feature classification	Rule-based scoring	Rule-based scoring	Rule-based scoring
Dependency	Performance measurement	Feature classification	Feature classification	Construct salience and message relevance
Data type	True/False	Score	Score	Score

13.2.4.5 Message tailoring process The message tailoring process has four steps that are designed to answer the following questions for each provider during the reporting period:

- 1. Identify performance features** What performance features are present or absent?
- 2. Infer construct salience** How salient are theoretical constructs as barriers to behavior change?
- 3. Assess component relevance** Which feedback message components are relevant?
- 4. Prioritize messages** What is the priority of each relevant feedback message?

The message tailoring process is initiated when report parameters, such as the individual provider or team identifier, reporting period, and performance indicators have been received. Each stage of the message tailoring process is described in Table 6. I describe each step of the tailoring process for one month of performance, for a single performance indicator and a single healthcare provider.

13.2.4.6 Identify performance features To conduct the feature assessment, the system identifies the presence or absence of performance features for the provider that are associated with the current month, such as the list of features described in section 13.2.3. These feature data are then stored in the performance database.

13.2.4.7 Infer construct salience To assess the salience of theoretical constructs, the system evaluates the tailoring rules described in section 13.2.4.1 using the features data collected for the provider's performance in the current month. Based on the presence and absence of features, the rules increase or decrease a score representing the salience of theoretical constructs, such as those discussed in section 13.2: Knowledge, social pressure, material resources, and self-efficacy. For the purpose of this research, I included the TDF domain of "skills" as a category of constructs for which general inferences could be made about the capability of individuals, using the following definition of skills: *"An ability or proficiency acquired through training and/or practice."*¹³⁶ An example of such an inference is that when individuals demonstrate consistently high performance for a behavior, the salience of all capability constructs as a barrier to behavior change, including skills constructs, is likely to be lower.

13.2.4.8 Assess component relevance To assess the relevance of feedback message components, the system uses tailoring rules such as those described in section 13.2.4.1. In the same approach used to estimate construct salience, these rules evaluate the presence and absence of performance features to score the relevance of message components for the provider's current month of performance. The rules estimate the relevance of feedback message components such as the following:

1. **Scale truncation:** The use of a truncated vertical axis to emphasize change in performance
2. **Self comparison:** Comparing an individual's past performance with current performance
3. **Peer comparison:** Comparing an individual's current performance with peer's current performance

4. **Historical peer comparison:** Comparing an individual's past performance with peer's past performance

After calculating the estimated component relevance scores for the current month, these data are written to the performance database.

13.2.4.9 Prioritize messages To assess the priority of tailored performance feedback messages, the system evaluates the both the message component relevance scores and estimated construct salience scores for the provider's current month of performance. The system uses rules to create priority scores for five feedback message types which correspond with the five different messages and tailoring approaches included in Figure 13. The rules that establish the message priorities represent theoretical causal mechanisms within the tailoring knowledge-base are the following:

1. If knowledge is salient as a barrier, then deprioritize withholding feedback
2. If skills are salient a barrier, then prioritize withholding feedback
3. If negative feedback has been delivered repeatedly, then prioritize withholding feedback
4. If self-efficacy is a barrier, then prioritize withholding feedback
5. If knowledge is a barrier, then prioritize current-score format
6. If skills are a barrier, then prioritize current-score format
7. If the salience of skills as a barrier is above 100, increase the priority of current-score format by salience-100
8. If self-comparison is relevant, increase the priority of self-comparison format
9. If skills are a barrier, then deprioritize peer-comparison
10. If peer pressure is a barrier then deprioritize peer-comparison
11. If peer pressure is a not barrier and peer-comparison is relevant, then prioritize peer-comparison
12. If self-efficacy is a barrier then deprioritize peer-comparison
13. If peer comparison and historical peer comparison are relevant, then increase the priority of historical peer comparison

After calculating the priority scores for each message type, for each individual and month of performance, these data are written to the performance database. After the conclusion of this assessment, a feedback tailoring system could generate a menu of tailored messages for a clinical supervisor to use for any individual who worked in any month during the two year period.

13.3 DISCUSSION

AF interventions can significantly impact the implementation of evidence-based practice. However, significant research effort in recent decades has been unable to answer the questions of how and when AF interventions will work.¹⁰ In response to Ivers and colleagues' call for new approaches to AF research¹¹, I argue that AF research should address a promising and novel AF component: automated feedback message tailoring systems. The potential significance of the systems I envision is growing with our increasing understanding of how to use eHealth data for comparative effectiveness research¹³ and with the development of standardized terminologies^{199,200,206} and common theoretical frameworks^{136,138} that create a basis for the use of computer-interpretable implementation knowledge. Furthermore, evidence about the use of computer-based message tailoring for health behavior change¹⁸⁴ and a significant understanding of knowledge-based computer systems in biomedical informatics¹⁶³ reveal a foundation of knowledge and tools that could support the development feedback tailoring systems. Perhaps most importantly, systems that provide support for the practice of giving performance feedback could create a helpful structure for clinical supervisors, who deal with much uncertainty and unanticipated reactions when giving feedback to health-care providers. I view this work as supporting a recognition of the complexity in providing evidence-based care that calls for improved judgment on the part of providers, rather than improved rule-following.²⁰⁷

The system architecture that I describe represents a new mode of AF that is adaptive and may potentially withstand the complexity of the clinical environment and individual differences in provider capability and motivation to improve feedback effectiveness. I have

outlined an approach to using knowledge representation methods to adaptively tailor feedback messages. A central part of this approach is to use the features of an individual's clinical behavior to make rule-based inferences about the causal mechanisms through which feedback influences future behavior.

13.3.1 LIMITATIONS

This research has several limitations. First, I have not evaluated the cost of development and maintenance of such a message tailoring system. It would seem that the use of message tailoring systems would be most cost-effective in larger health systems where eHealth is already used to support performance measurement, but the cost-effectiveness of this approach is an important area of future research.

While I believe the rule-based approach to modeling construct salience that I used was adequate for the purpose of an exploratory analysis, it is likely to be inadequate for a large number of rules or to model the complexity of additional constructs. Using a Bayesian network to probabilistically model the network of factors influencing feedback effectiveness is likely to be a more viable approach. The benefits of such an approach have been discussed in the context of intelligent tutoring systems.²⁰⁸ The primary benefits of using a probabilistic approach are that it can adequately represent complex interactions resulting from multiple observations, and it can allow for the explicit representation of supervisors' beliefs about feedback recipient's barriers to behavior change. These beliefs could be modeled and revised over time as supervisors observe the effect of repeated feedback on individual performance, and change their beliefs about the effectiveness of feedback message designs for individual providers. This network would require a "recipient model" that could probabilistically represent the feedback recipient's capability, opportunity, and motivation factors with regard to each performance indicator. Additionally, the network could represent the recipient's reported preferences for receiving feedback to estimate the probability that a tailored feedback message in a menu would lead to improved performance.

Another limitation is that the system's ability to provide effective feedback is contingent on the ability of a supervisor to accurately perceive specific barriers for each individ-

ual. Supervisors' ability to identify barriers can be expected to vary across supervisors and situations, and could contribute to the ineffectiveness of feedback. However, I note that, compared to feedback which is not tailored for specific barriers, we can reasonably expect that a message tailoring tool could provide relative improvement to the effect of feedback messages. Nevertheless, I do not know the extent to which making inaccurate assumptions about barriers to behavior change could negatively impact performance. For the purpose of our examples, I did not validate the tailoring rules that represent the relationship between a theoretical causal mechanism and the performance features found in the data.

13.4 CONCLUSION

Understanding how to tailor feedback messages holds significant potential for the improvement of AF interventions. In pursuing the goal of understanding how to develop tools for automated feedback tailoring, I plan to evaluate a prototype feedback message tailoring system in disparate AF intervention settings. This work is perhaps best characterized as embracing the complexity of healthcare by developing adaptive tools to target individual providers' specific barriers to the adoption of evidence-based practice.

14.0 AUTOMATED FEEDBACK TAILORING IN LOW-RESOURCE SETTINGS

The final study I conducted for this dissertation brought together the methods, knowledge, and tools developed in the previous three studies. Having designed a system architecture to support automated feedback tailoring adaptively in complex healthcare settings, I aimed to evaluate the system design in a low-resource setting. The purpose of this study was to evaluate the potential for a feedback message tailoring system to impact clinical performance in ART clinics in Malawi. I chose the following two distinct objectives for this research: Firstly, to formatively evaluate the design of the automated feedback tailoring system, and secondly to understand the potential impact of such a system in ART clinics in Malawi. I approached the analysis of potential impact in two further sub-aims. These were firstly to identify clinical behaviors that were measurable and had improvement potential, and secondly to identify “room for tailoring” (e.g. to identify barriers to behavior change that were variable and justify the need for tailoring, as opposed to a standardized approach).

14.1 METHODS

To conduct this research, I collected de-identified EMR data from ART clinics in public hospitals in Malawi. I analyzed the EMR data using guideline-based performance indicators and the prototype tailoring knowledge base that I developed in Chapters 11-13. This research was approved by the University of Pittsburgh Institutional Review Board (IRB), protocol PRO12100159 and the Malawi National Health Sciences Research Committee (NHSRC) protocol #1019.

Data collection for this study occurred in November, 2013. I collected two years of de-identified EMR data from October, 2011 to September, 2013, from 11 ART clinics in Malawi that were using the Malawi National ART EMR (Software version BART 1). The data was de-identified and made available to me by Baobab Health Trust, in Lilongwe, Malawi, where the organization maintains and develops the National ART EMR. I conducted data analyses for this research in three phases, each designed to answer the following questions at the level of the individual healthcare provider:

Performance measurement Can we use EMR data to measure performance?

Message tailoring and prioritization Which tailored feedback messages have the highest priority?

Potential impact of tailoring How frequently could feedback tailoring impact performance?

14.1.1 Performance measurement

The purpose of this analysis was to understand if there are opportunities to provide performance feedback in ART clinics in Malawi by measuring performance using EMR data. If no credible performance information can be extracted from the EMR data, then no opportunities to provide feedback exist, therefore this was a first critical step. In the study I conducted in Chapter 11, I found that a set of 21 recommendations were auditable, based on EMR data from four ART clinics in Malawi, therefore I anticipated that performance measurement would be feasible. However, given significant revisions to Malawi's ART guideline and the EMR software which had occurred between the prior research and this study, I was unable to use the earlier set of auditable recommendations. Based on my first-hand experience of change in guideline recommendation auditability, I sought to identify stable performance indicators for a range of clinical behaviors, where performance is defined as compliance with desired clinical practice, at the individual healthcare provider level.

I selected four standard performance measures for ART treatment that have been used in multiple Sub-Saharan African countries to improve the quality of care.¹⁹⁴ These indicators are aligned with guideline statements from Malawi's national guideline for the clinical management of HIV, 2011 edition (Table 7), and have not changed across the 2008 and 2011

versions of Malawi's ART guidelines. Each measure contains a numerator and denominator that is used to calculate a percentage of adherence to recommended clinical practice.

Table 7: Four performance indicators mapped to Malawi's national ART guideline.

Performance indicator	Malawi ART guideline recommendation	Numerator	Denominator
Monitoring of nutritional status: Pediatric patient height	"Record length / height to the nearest cm at every visit (children)" (2011 edition, page 18)	Number of children with height recorded at least once during the review period	Number of children with at least one clinical visit during the review period
Monitoring of nutritional status: Weight	"Record weight in kg to the nearest 100g at every visit" (2011 edition, page 18)	Number of patients with weight recorded at least once during the review period	Number of patients with at least one clinical visit during the review period
CPT prescribing	"Provide CPT to all patients in HCC and ART follow-up" (2011 edition, page 32)	Number of patients who were prescribed CPT	Number of patients with at least one clinical visit during the review period who did not have CPT contraindications
WHO Staging	WHO clinical staging is mandatory for all HIV patients, including those who are universally eligible for ART (confirmed infected children under two years, pregnant or breastfeeding women) or those with a CD4 count result (2011 edition, page 12)	Number of patients with a WHO clinical stage at the time of ART initiation	Number of patients who were initiated on ART during the review period

The performance indicators that I identified indicate possible performance problems, but not actual performance problems. Performance indicators are commonly used to identify possible problems that may in fact represent data quality problems or valid exceptions to recommended clinical practice.²⁰⁹ I measured performance as a precursor to the identification of data quality problems that performance feedback could potentially be used to address, in addition to addressing actual performance problems.

Performance for each indicator is calculated from the ART EMR data using the methods described in Chapter 11. The standard performance indicators I used are typically calculated at a clinic level, but for this research I adapted the indicators to measure performance at the level of the individual provider. The denominator reflects the total number of opportunities a provider had to provide recommended care to each patient. For example, if a provider was recorded as conducting an ART visit with a patient who was eligible to receive CPT, this patient was counted towards the total number of patients in the denominator for the month of that visit. The numerator reflects the documented care received by the patients who were counted in the individual provider’s denominator, regardless of who provided the care to the patient. In this way the indicators are a lower-bound for performance, and do not penalize providers unfairly for team-based care. For example, in a scenario where a provider does not prescribe CPT to an eligible patient at the time of an ART visit, but the patient receives a prescription for CPT on another day that month, from any other provider, the patient would still be counted in the first provider’s numerator. An exception to this lower-bound is for patients whose care happens to be provided adequately but at a time frame spanning the end of one month and the start of the next month.

Using the Ruby programming language, MySQL, and R, I created scripts that measured and displayed performance for each of the four performance indicators. To validate the results I reviewed the queries, scripts and the performance data with EMR developers from Baobab Health Trust.

14.1.2 Message tailoring and prioritization

After measuring performance for all providers over the two-year period, I used the prototype feedback message tailoring system to analyze the performance data for each indicator to prioritize a set of tailored feedback messages for all providers and all months, using the approach described in section 13.2.4.5. I adapted the preliminary message tailoring knowledge base developed in Chapter 13 with the total number of rules shown in Table 8. The message tailoring process generated five message types that are featured in the menu of tailored messages (Figure 13) from section 13.2. I labeled the five message types from Figure 13,

beginning with the first row, as follows: *withhold feedback*, *current score*, *self comparison*, *peer comparison*, and *historical peer comparison*. I created two additional categories for prioritization: a) *no prioritization* for the cases where no message types could be prioritized, and b) *prioritized combination*, for the cases where two or more messages were equally of highest priority.

Table 8: Totals of message tailoring rules used in the message tailoring process.

	Performance feature classification rules	Construct salience inference rules	Message relevance inference rules	Prioritization rules
Total	11	5	6	13

14.1.3 Potential impact of tailoring

Building on the performance measurement and feedback tailoring assessments using EMR data from ART clinics in Malawi, I evaluated “room for tailoring” by identifying opportunities to provide different types of feedback messages, and observing how frequently message tailoring would yield differences in message priorities for this setting. To evaluate these opportunities and message differences I conducted two analyses designed to answer the following questions:

Performance gaps Where are there gaps in performance between peers at a site?

Message variability How variable is the priority of feedback messages?

14.1.3.1 Performance gaps Peer comparison feedback is widely used for quality improvement to encourage low-performing providers to change clinical behaviors. To understand how frequently peer comparison feedback could be provided, I calculated the proportion of months between October, 2011 and September, 2013 that had one or more performance gaps. I defined a performance gap as a 10% difference in performance between an individual healthcare provider and the average performance of two peers working in the same clinic in the same month. To minimize the influence of healthcare providers who saw few patients

on the frequency of performance gaps, I excluded providers who had treated 10 or fewer patients in each month. For each performance indicator, I calculated the percentage of monthly reports containing performance gaps that are greater than 10% between at least one individual and the monthly average of two of their high-performing peers. I calculated summary statistics for this percentage across all 11 sites, for two years of monthly report data, from October 2011 to September 2013. I also calculated the monthly total of performance gaps, and summary statistics for this total during the reporting period.

14.1.3.2 Message variability Understanding the variability of message priorities allows system stakeholders to determine the degree to which message tailoring is useful, relative to the delivery of a standardized feedback format for all providers. When there is increased variability of message priorities, tailoring will be more useful because feedback is likely to be improved by accommodating individual and situational differences among healthcare providers. If for example, the system generates messages having the same priority for more than 95% of providers, it would suggest that message tailoring is not necessary, because the potential impact of message tailoring will be low. However, if the priority of messages is more evenly stratified across message type groups, and if the size of these groups changes over time, it would suggest a greater potential impact for automated message tailoring.

For each performance measure across all individuals and months during the reporting period, I created a prioritized list of message types. To assess the variability of message priority, I calculated the percentage of individual performances that had each of the five types of message as the highest priority, plus the two additional prioritization categories. To further assess the degree to which any stratification of providers across message groups changed between months, I calculated the absolute change in the percentage of each type of highest priority message over all months.

14.2 RESULTS

14.2.1 Performance measurement

For each of the four performance indicators, I calculated individual performance for a total of 372 unique healthcare providers at 11 ART clinics who worked during the two year period. I calculated performance for the four indicators at a daily frequency, for a total of 73,185 daily performance reports with a denominator of one or more patients. I summarized daily reports at a monthly frequency by individual provider for a total of 7,448 individual monthly performance reports with a denominator of five or more patients (Table 9). I found that on average per performance indicator, per site, and per month approximately seven automated monthly performance reports could be generated. In district hospitals this average was approximately 5.1 reports, while in central hospitals, it was 15.9 reports.

Table 9: Individual monthly report totals summary statistics.

Indicator	Individual monthly reports	Reports generated per site, per month			
		M	SD	Min	Max
Pediatric height recording	1,193	4.5	2.90	1	20
Weight recording	2,506	9.5	6.24	1	33
WHO clinical staging	1,197	4.6	2.97	1	18
CPT prescribing	2,552	9.7	7.78	1	37
Total	7,448	7.1	5.96	1	37

Summary statistics about the performance of healthcare providers are shown in Table 10. The mean performance ranged from 69% for pediatric height recording ($SD = 0.347$) to 97% for WHO clinical staging ($SD = 0.042$).

Table 10: Mean performance of 11 ART clinics in Malawi, 10/2011 to 9/2013.

Indicator	Performance	
	M	SD
Pediatric height recording	69%	0.347
Weight recording	96%	0.051
WHO clinical staging	97%	0.042
CPT prescribing	73%	0.355

Summary statistics for each of the four guideline-based performance measures for all healthcare providers grouped by ART clinic are shown in Table 11. The average monthly performance for weight recording and WHO clinical staging were consistently high ($> 94\%$) for all but one clinic for each indicator. The average monthly performance for pediatric height recording ranged from 2.3% to 98.4%, while the average monthly performance for CPT prescribing ranged from 48.9% to 87.6%.

The monthly performance of 11 ART clinics in Malawi for each of the four performance indicators between October, 2011 and September, 2013 is shown in Figures 15, 16, 17, and 18. Monthly performance of weight recording (Figure 16) and WHO clinical staging (17) show relatively consistent, high performance for the last 12 months of the period, with the exception of weight recording for providers at Clinic five. Performance of pediatric height recording (Figure 15) and CPT prescribing(18) appears more variable. For pediatric height recording, providers at four clinics (2, 5, 8, and 9) are consistently below the performance of the other clinics. For CPT prescribing performance, there is a generalized decrease in performance in 2012 across clinics, with providers at most clinics having a wide range of performance during the period. The generalized decrease in performance is associated with a national shortage of CPT drugs that occurred in 2012.

Table 11: Mean performance of 11 ART clinics in Malawi for four indicators.

Clinic	Pediatric height recording		Weight recording		ART staging		CPT prescribing	
	M	SD	M	SD	M	SD	M	SD
1	91.0%	0.065	95.1%	0.031	98.2%	0.032	81.1%	0.294
2	48.3%	0.159	94.9%	0.035	96.5%	0.042	87.6%	0.220
3	98.3%	0.021	99.4%	0.008	98.3%	0.017	83.3%	0.303
4	96.3%	0.037	99.6%	0.003	99.3%	0.010	65.2%	0.424
5	23.7%	0.138	84.3%	0.077	98.1%	0.021	60.0%	0.417
6	82.3%	0.079	95.1%	0.018	99.4%	0.011	65.5%	0.415
7	92.5%	0.029	98.5%	0.003	89.4%	0.074	74.0%	0.270
8	37.7%	0.227	98.3%	0.008	98.2%	0.020	81.1%	0.328
9	2.3%	0.017	99.3%	0.006	95.5%	0.036	48.9%	0.309
10	91.0%	0.063	97.5%	0.015	98.3%	0.030	73.3%	0.390
11	98.4%	0.009	99.1%	0.005	97.1%	0.017	79.7%	0.263

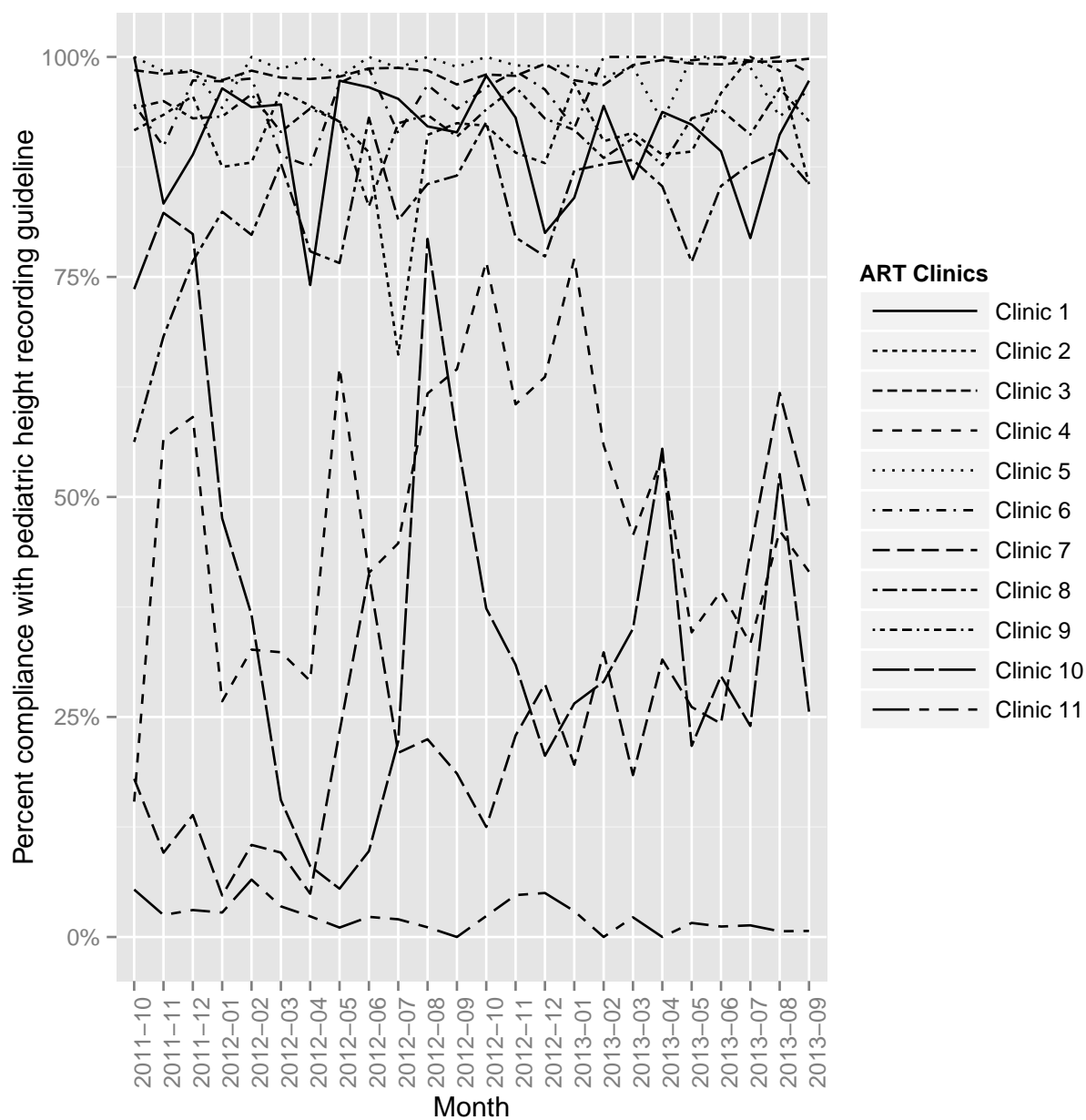


Figure 15: Monthly pediatric height recording performance, 10/2011 to 9/2013.

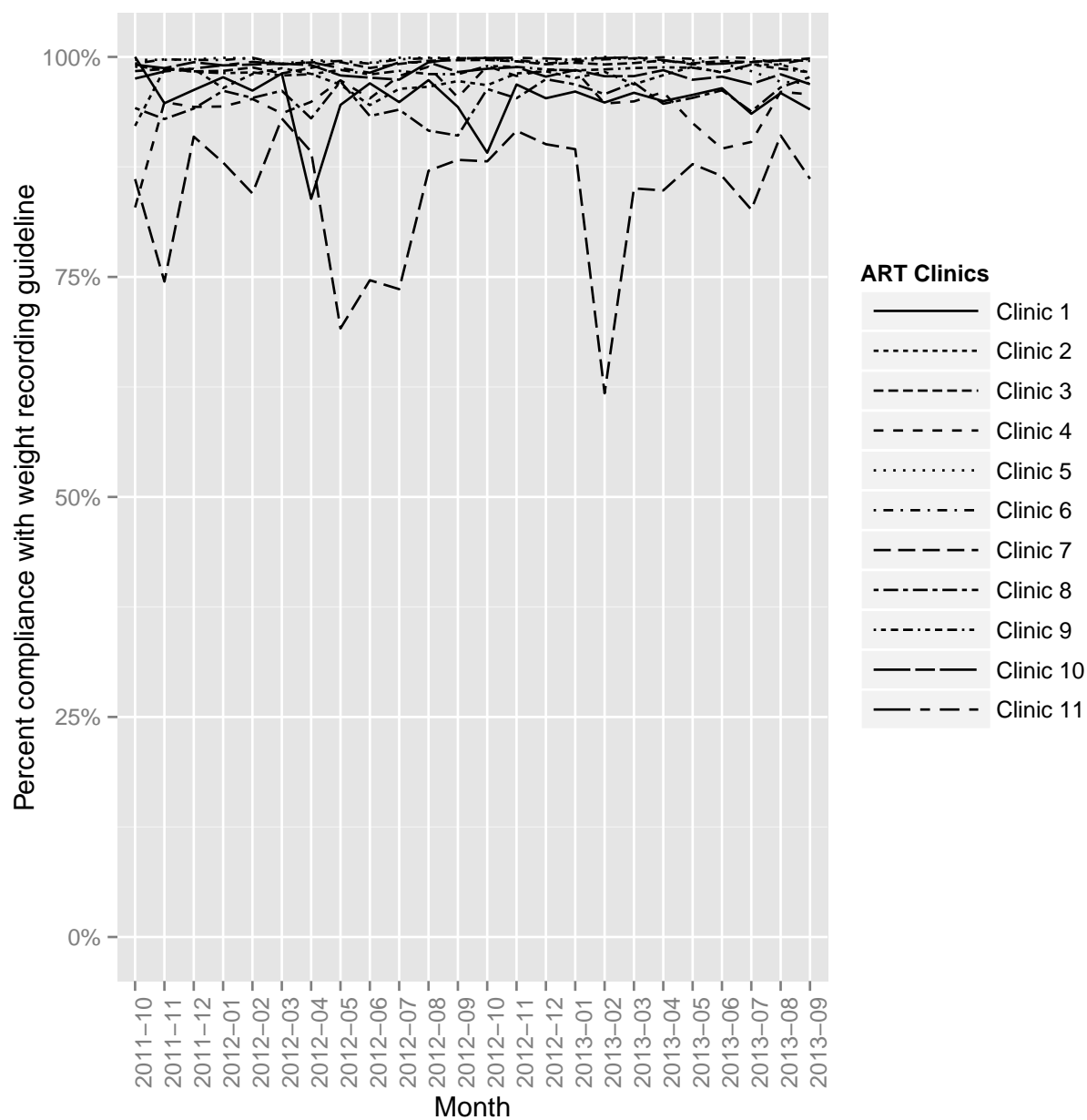


Figure 16: Monthly weight recording performance, 10/2011 to 9/2013.

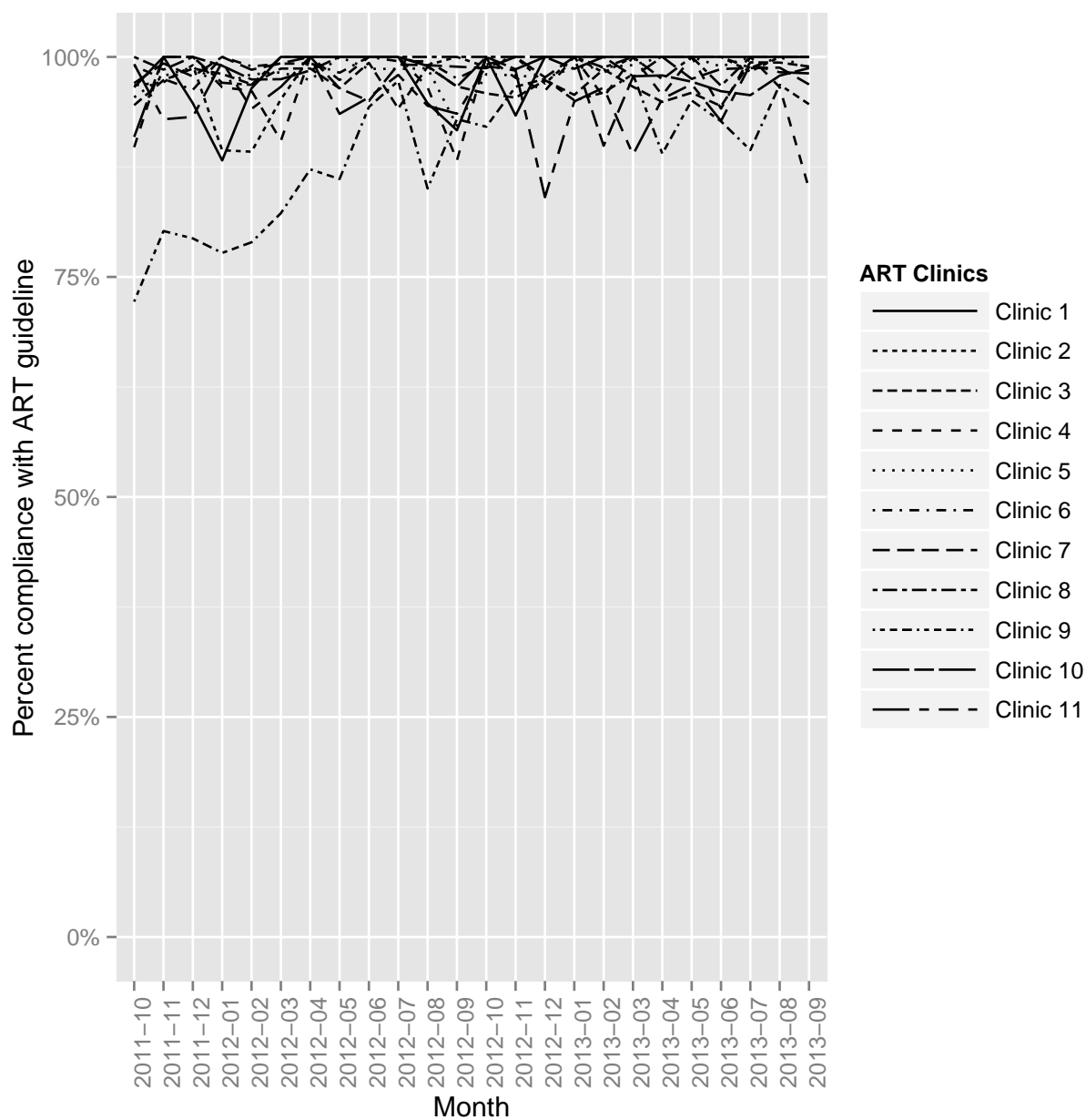


Figure 17: Monthly WHO clinical staging performance, 10/2011 to 9/2013.

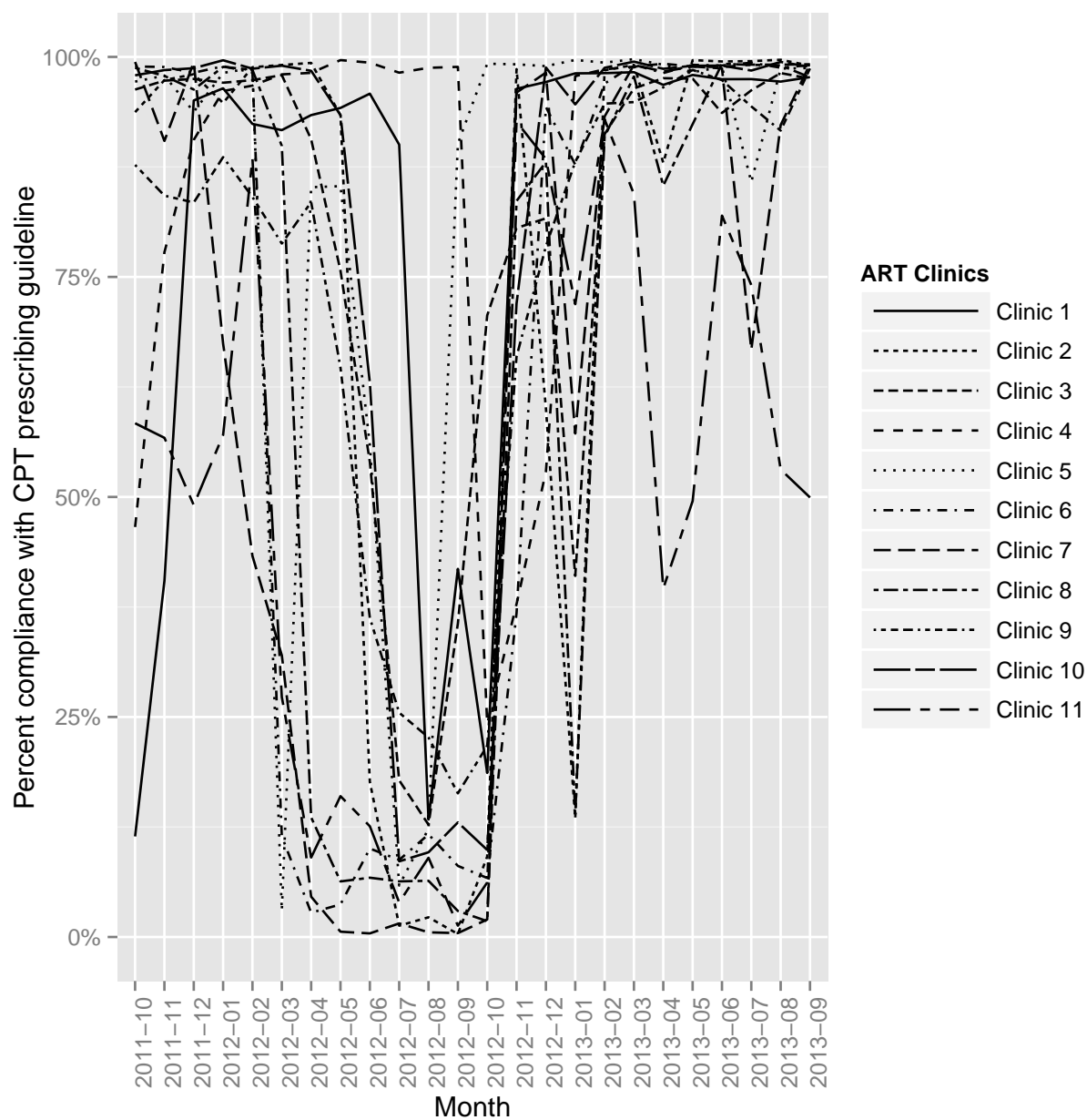


Figure 18: Monthly CPT prescribing performance, 10/2011 to 9/2013.

14.2.2 Feedback message tailoring

The characteristics of the data I collected and analyzed for the feedback tailoring steps are shown in Table 12.

Table 12: Characteristics of data collected and analyzed for feedback message tailoring.

Indicator	Individual monthly reports	Feature observations (N = 11)	Saliency scores (N = 5)	Relevance scores (N = 4)	Priority scores (N = 5)
Pediatric height recording	1,193	13,172	5,965	4,772	5,965
Weight recording	2,506	27,585	12,530	10,024	12,530
WHO clinical staging	1,197	13,167	5,985	4,788	5,985
CPT prescribing	2,552	28,072	12,760	10,208	12,760
Total	7,448	81,928	37,240	29,792	37,240

14.2.3 Potential impact of tailoring

14.2.3.1 Performance gaps I calculated the monthly performance gap total for each performance indicator across 11 ART clinics between October, 2011 and September, 2013. Summary statistics for the performance gaps are shown in Table 13. The mean percentage of months with performance gaps ranged from 11% (SD = 0.189) for WHO clinical staging to 56% (SD = 0.351) for weight recording. Notably, for WHO clinical staging, a majority (6 out of 11) of the clinics had no performance gaps at all during the two years analyzed. In contrast to WHO clinical staging, a majority of clinics for both weight recording and CPT prescribing had performance gaps in more than 1/3 of the months analyzed during the two years.

Table 13: Percentage of months containing performance gaps at 11 ART clinics in Malawi.

Clinic	Pediatric height recording		Weight recording		ART staging		CPT prescribing	
	%	N	%	N	%	N	%	N
1	0%	0	29%	7	0%	0	13%	3
2	4%	1	75%	18	0%	0	46%	11
3	8%	2	4%	1	8%	2	25%	6
4	25%	6	13%	3	0%	0	50%	12
5	8%	2	96%	23	0%	0	42%	10
6	42%	10	96%	23	0%	0	42%	10
7	54%	13	100%	24	63%	15	96%	23
8	75%	18	42%	10	0%	0	33%	8
9	8%	2	46%	11	25%	6	88%	21
10	33%	8	33%	8	8%	2	46%	11
11	17%	4	83%	20	13%	3	100%	24
Mean	25%	6.0	56%	13.5	11%	2.5	53%	12.6
(SD)	(0.238)	(5.71)	(0.351)	(8.43)	(0.189)	(4.55)	(0.290)	(6.96)

The mean monthly total of performance gaps for each indicator across sites between October, 2011 and September, 2013 is shown in Figure 19. The average number of performance gaps that could be used to give peer comparison feedback to a single provider for all 11 sites ranged from 0.32 to 2.45 gaps per month.

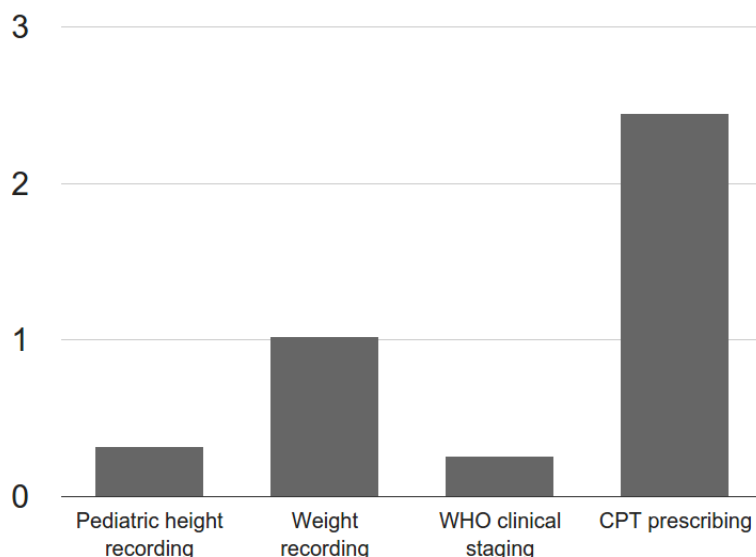


Figure 19: Mean monthly performance gap totals, 10/2011 to 9/2013.

The monthly performance gap totals are shown in Figures 20, 21, 22, and 23. The monthly performance gap totals were consistently low (three or lower) for pediatric height recording (20) and WHO clinical staging (22) with the exception of Clinic 7 prior to July of 2012. Weight recording performance had slightly higher monthly performance gap totals, with an average total of slightly above one performance gap per month, per clinic ($M=1.02$, $SD=1.20$). CPT prescribing had the highest average total of gaps ($M=2.45$, $SD = 4.06$) and the greatest variability in gap totals. Clinics 7 and 11 stood out from other clinics in having more gaps overall, averaging around eight gaps per month each for CPT prescribing. These clinics were the only two central hospital ART clinics, which had the highest provider total and therefore have an increased potential for performance gaps to occur between providers.

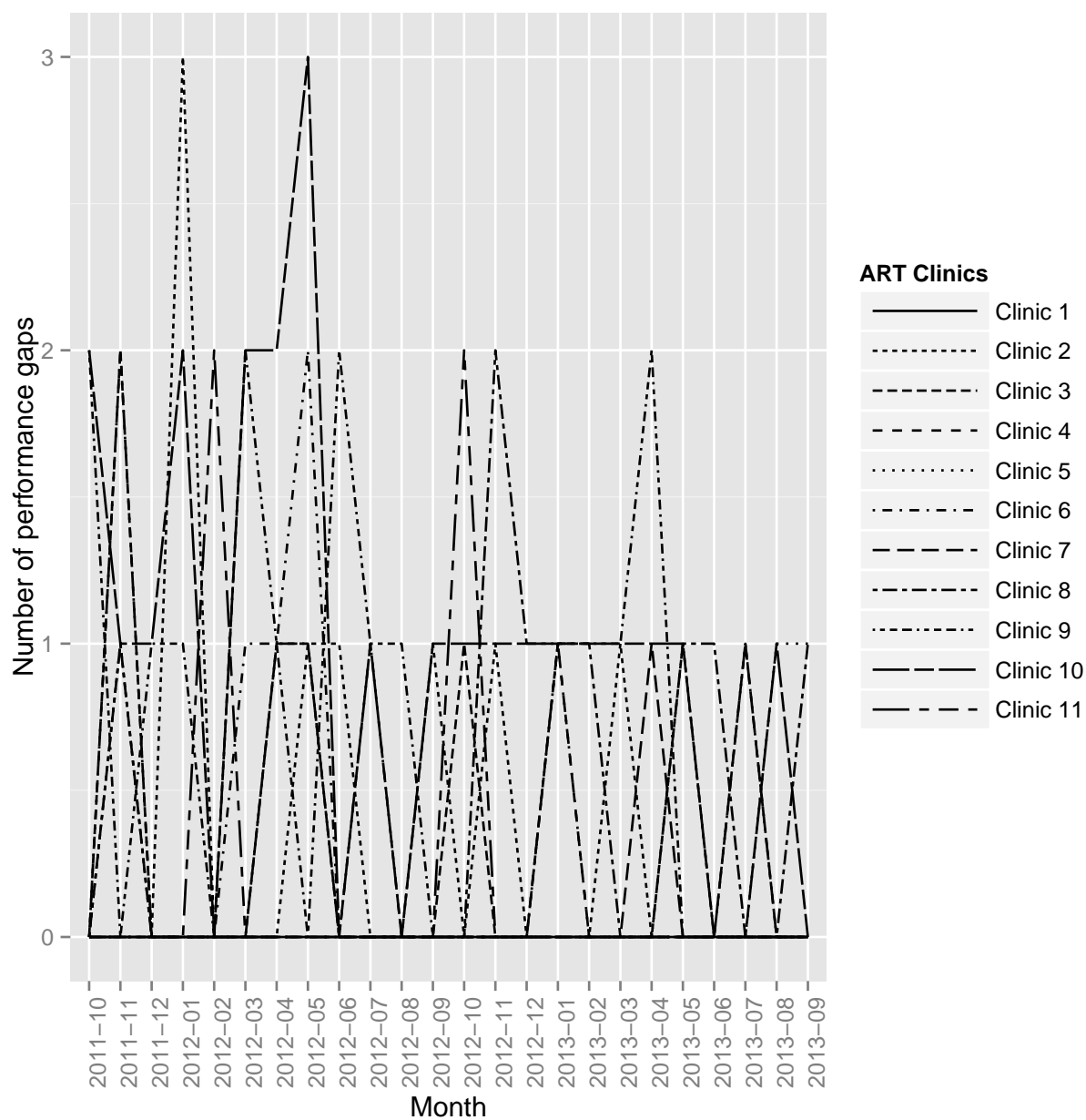


Figure 20: Performance gap totals for pediatric height recording, 10/2011 to 9/2013.

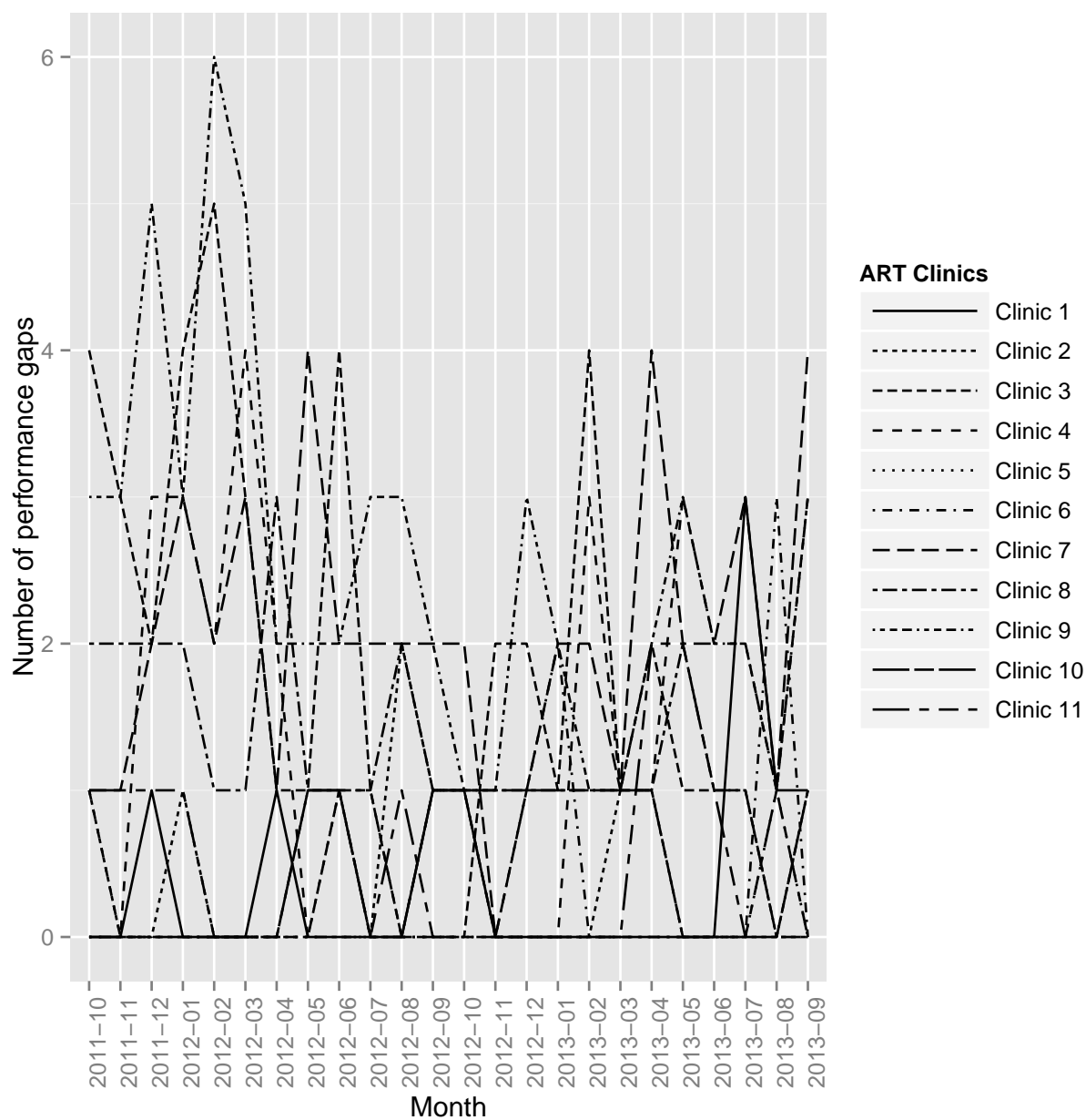


Figure 21: Performance gap totals for weight recording, 10/2011 to 9/2013.

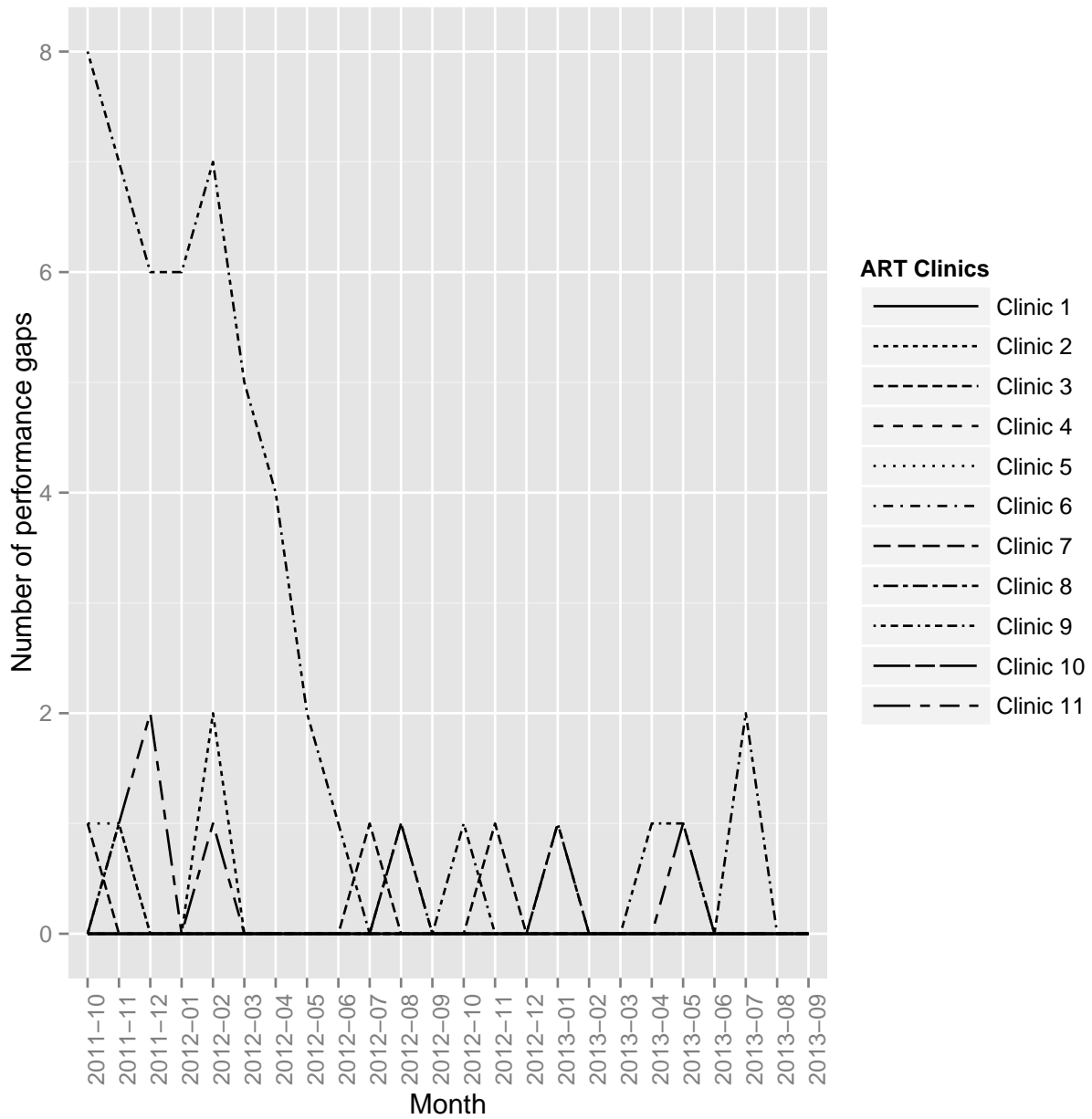


Figure 22: Performance gap totals for WHO clinical staging, 10/2011 to 9/2013.

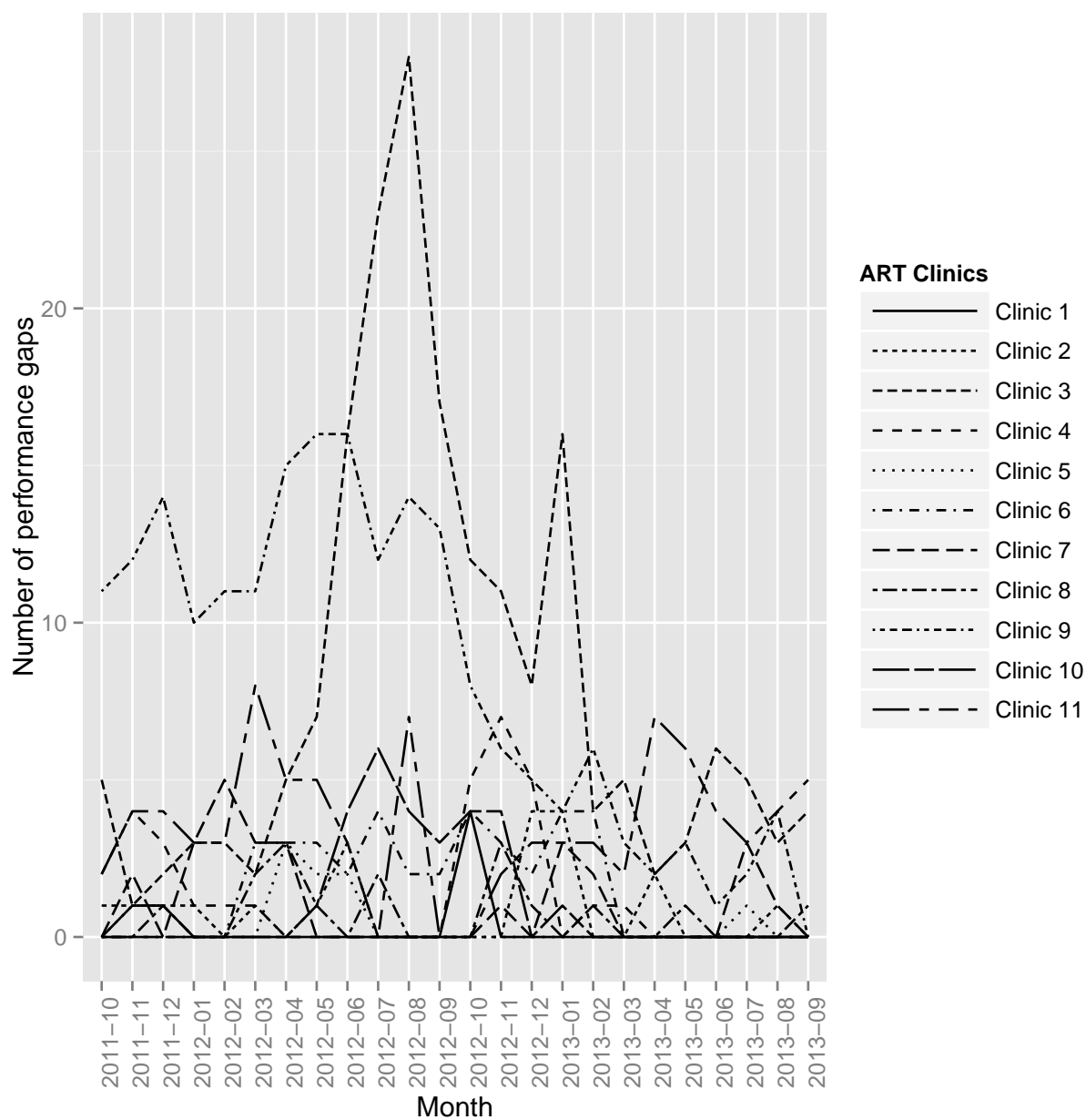


Figure 23: Performance gap totals for CPT prescribing, 10/2011 to 9/2013.

14.2.3.2 Message variability The tailoring process resulted in 35% (2,624 / 7,448) of individual monthly reports being prioritized to optimize the effect of feedback on performance. I calculated the percentage of all messages that had each message type as the highest priority on an individual's monthly report. No reports had *peer comparison* or *historical peer comparison* messages as the highest priority. The mean percentages of each message type for all clinic's messages, for each performance indicator are shown in Figure 24. Across all performance indicators, increased stratification of tailored message types appears to be associated with lower performance. For example, the indicators having higher performance, which are weight recording and WHO clinical staging, had a higher average percentage of messages that were not prioritized, at 75% for weight and 83% for WHO clinical staging. In contrast, pediatric height recording and CPT prescribing, which have lower overall performance, had increased stratification of highest priority percentages across message types.

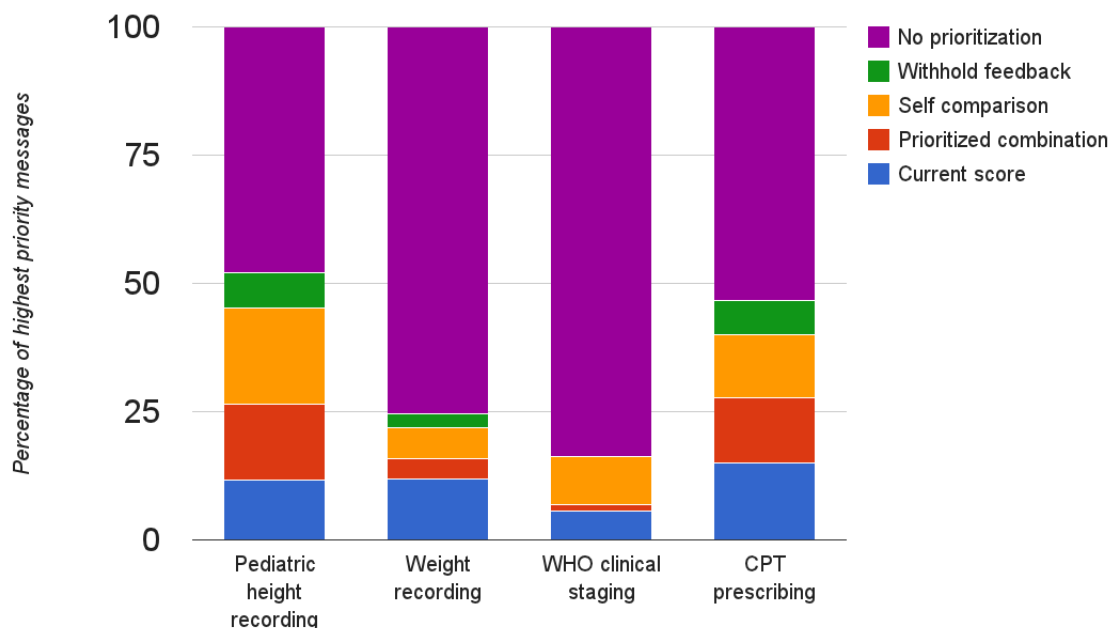


Figure 24: Mean percentages of prioritized message types.

The mean percentage of tailored messages that were of the highest priority for individual providers are shown in Figures 25, 26, 27, and 28. These figures show the monthly change in the average percentage of highest-priority message types.

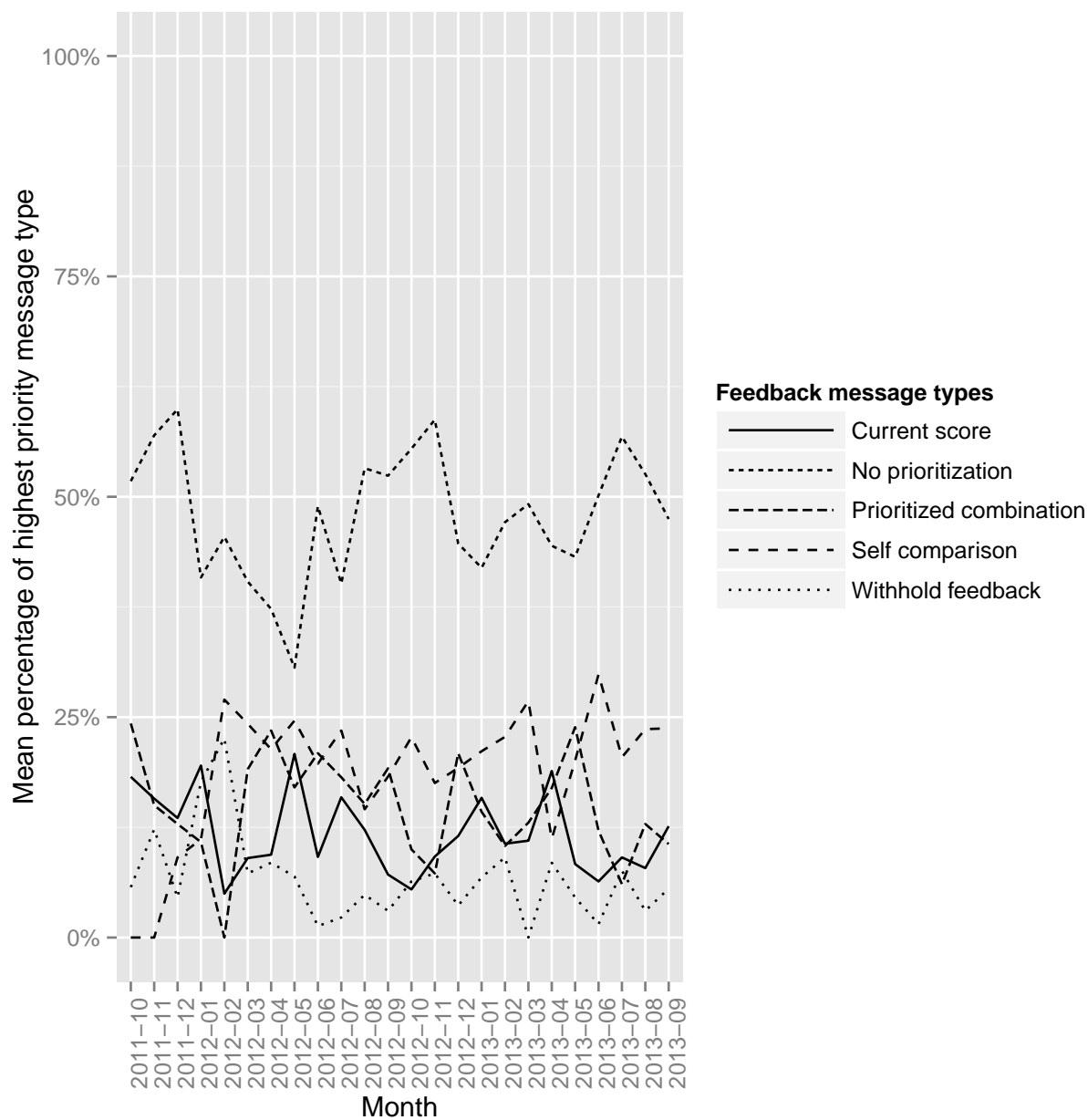


Figure 25: Mean percentage of tailored messages prioritized for pediatric height recording.

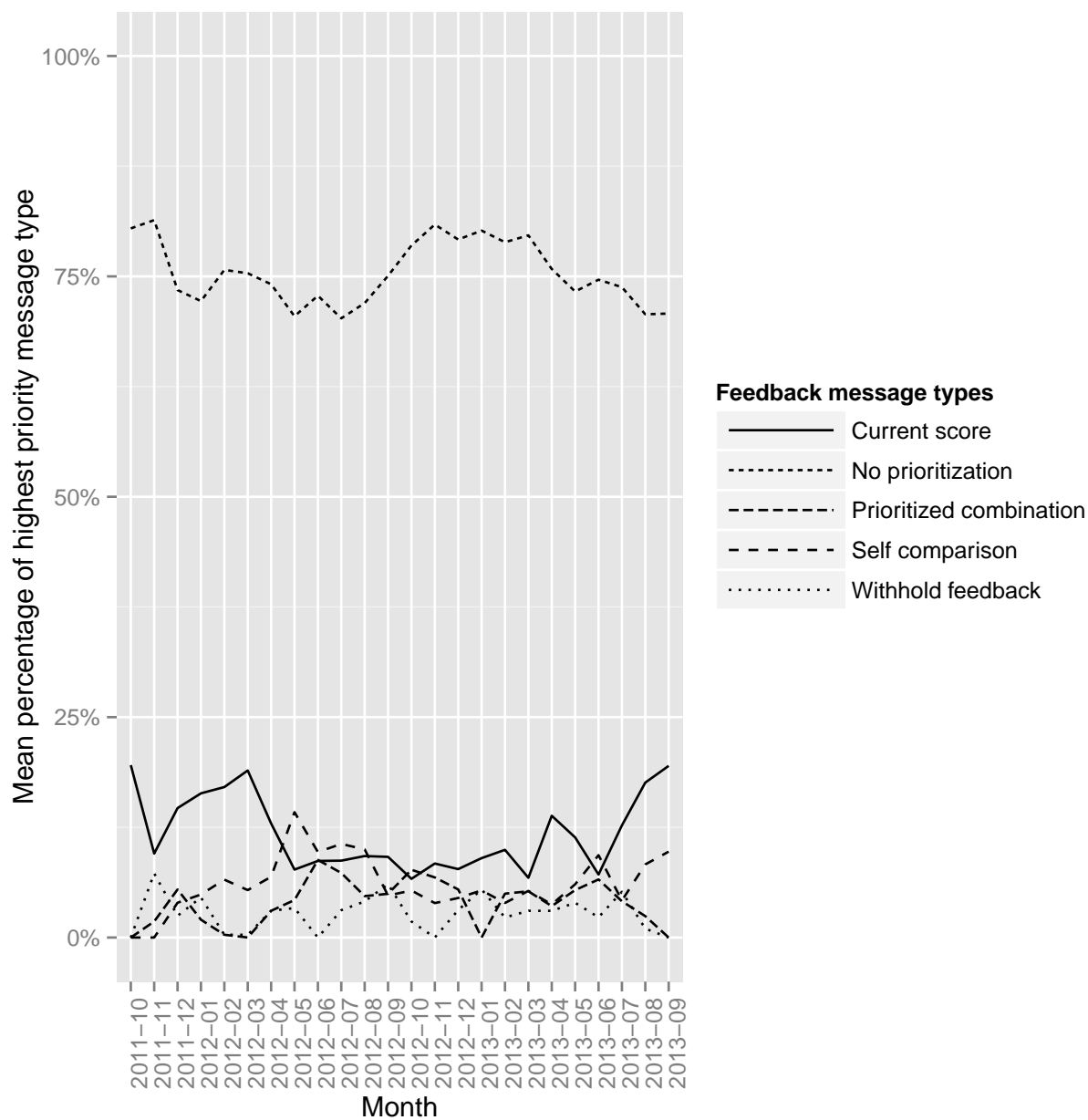


Figure 26: Mean percentage of tailored messages prioritized for patient weight recording.

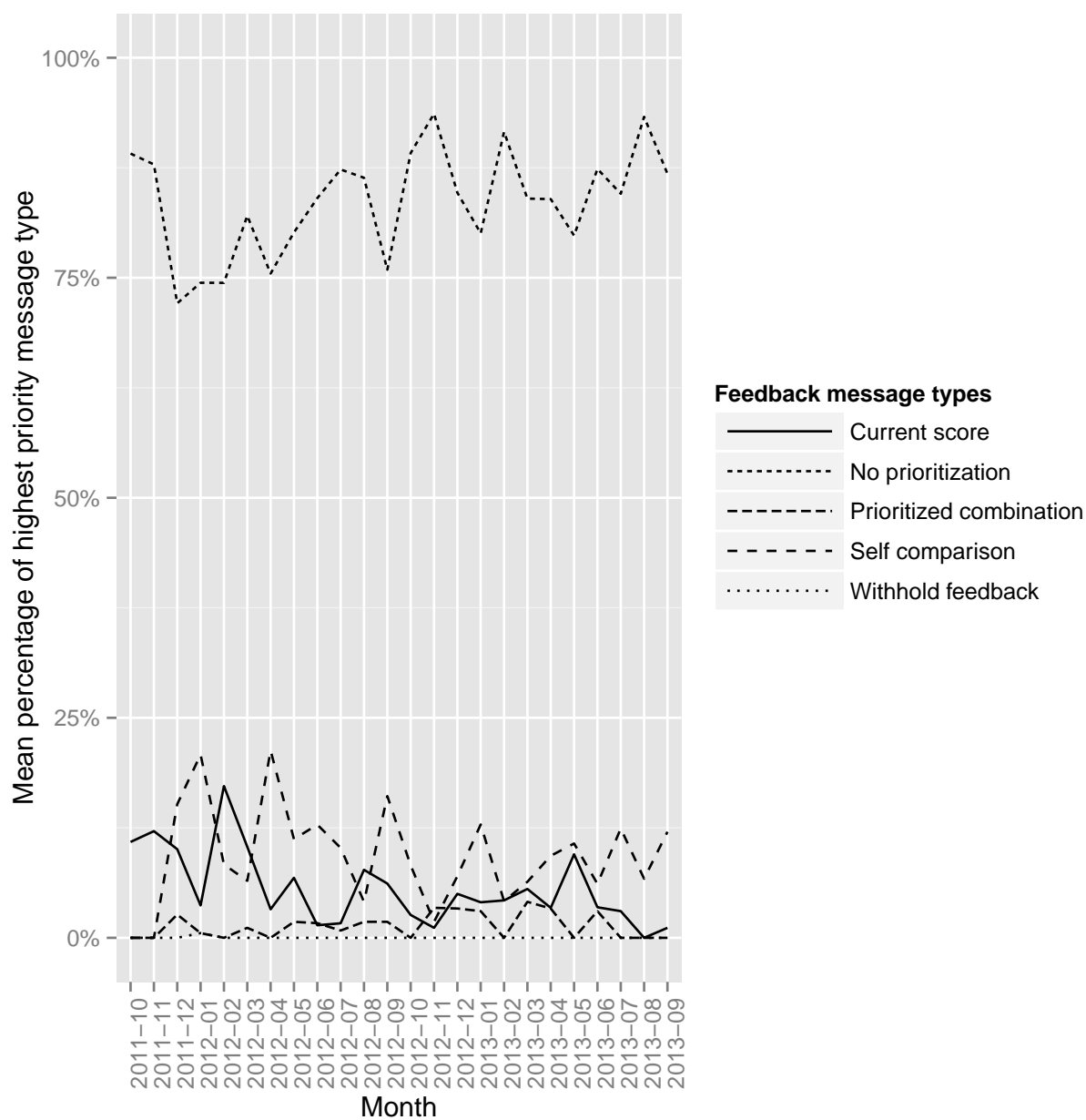


Figure 27: Mean percentage of tailored messages prioritized for WHO clinical staging.

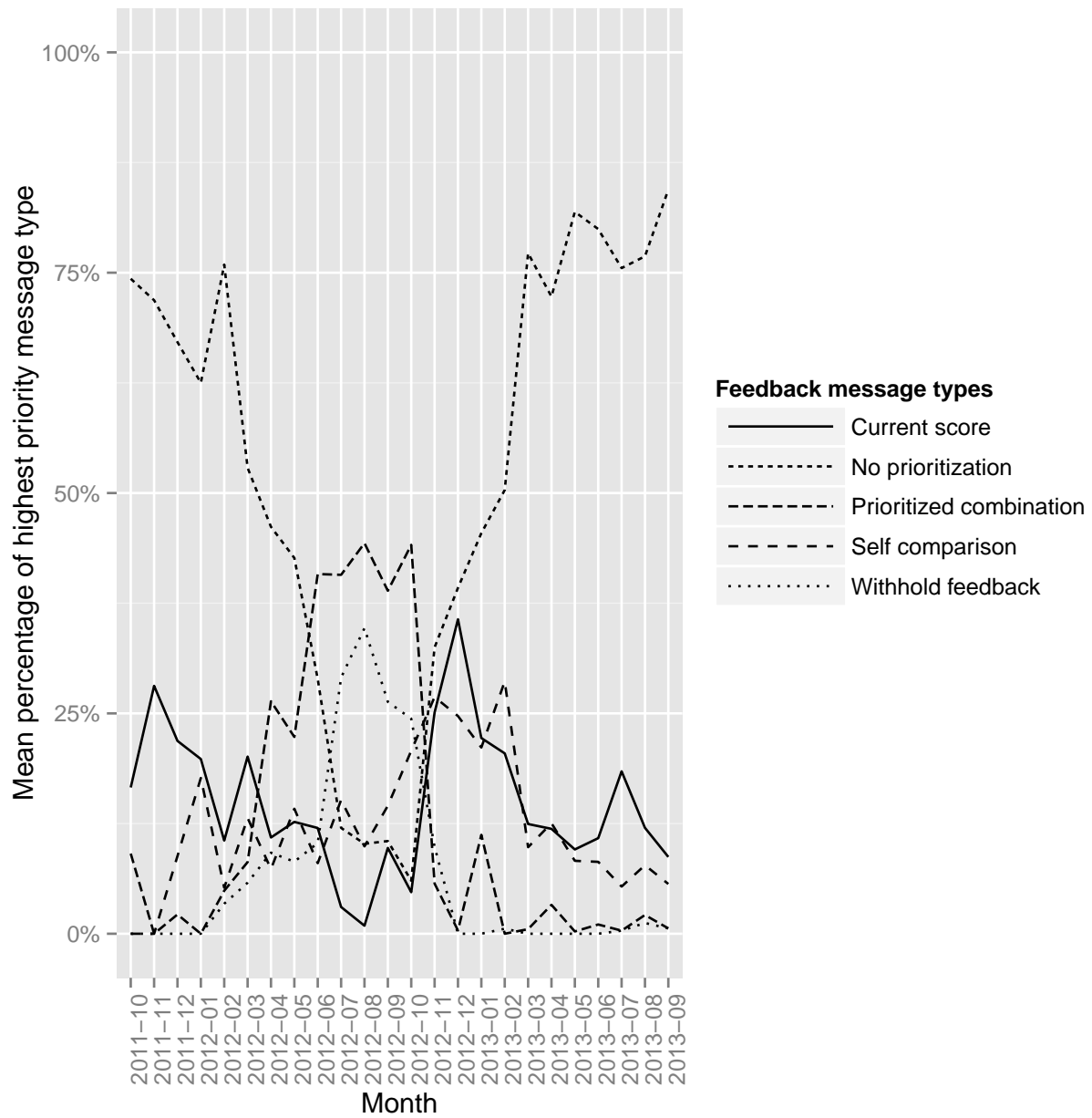


Figure 28: Mean percentage of tailored messages prioritized for CPT prescribing.

To further assess the degree to which any stratification of prioritized messages across message types changed between months, I calculated the mean absolute change in the percentage of each type of highest priority message over all months (Table 14).

Table 14: Mean monthly absolute differences in percentage of prioritized message types.

Indicator	Absolute difference	
	M	SD
Pediatric height recording	13.0%	0.203
Weight recording	6.6%	0.104
WHO clinical staging	6.8%	0.146
CPT prescribing	12.6%	0.193

The mean absolute percentage change in the proportions of tailored messages for 11 ART clinics between October, 2011 and September, 2013 are shown in Figures 29, 30, 31, and 32.

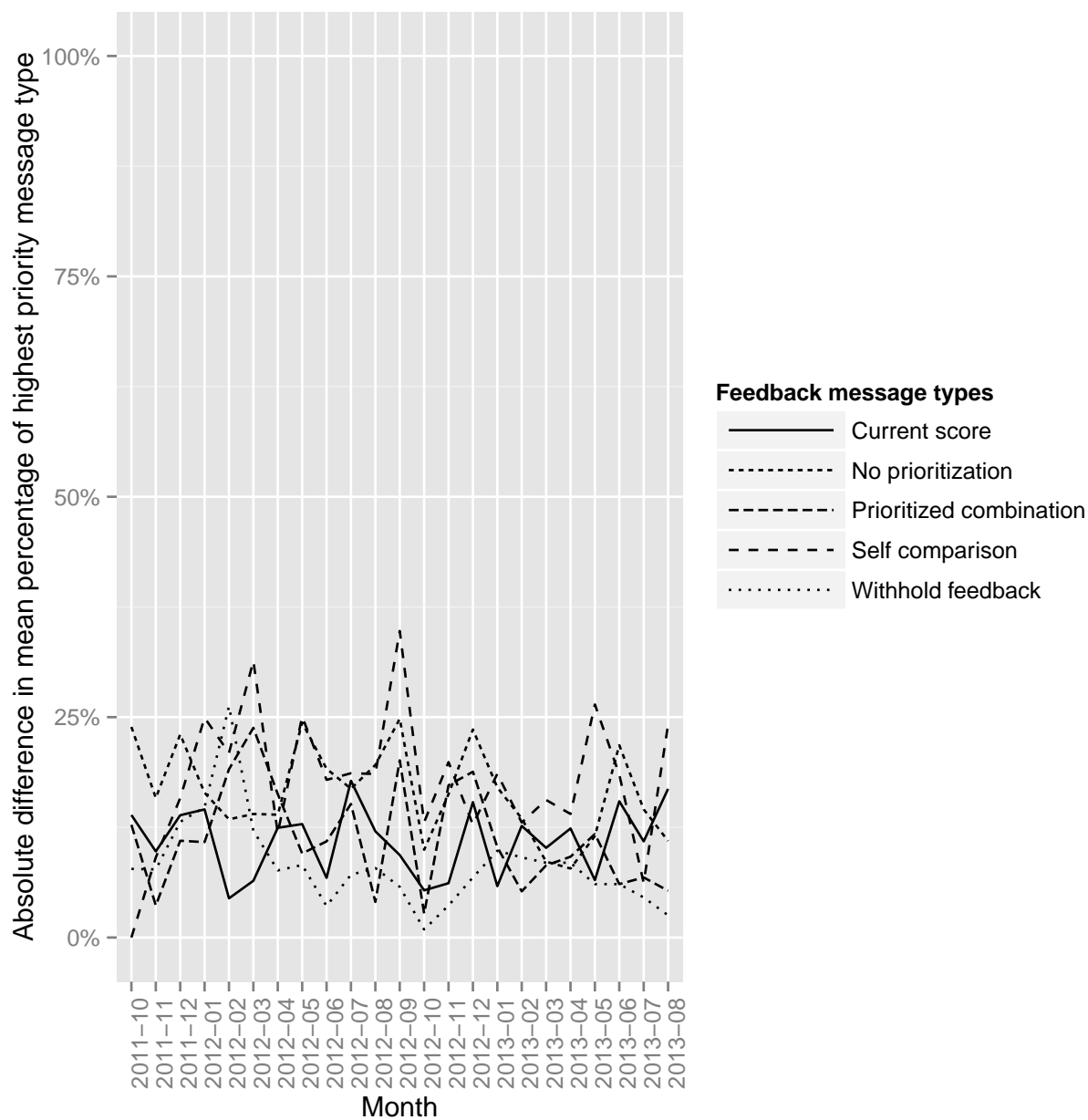


Figure 29: Mean absolute percentage change in messages for pediatric height recording.

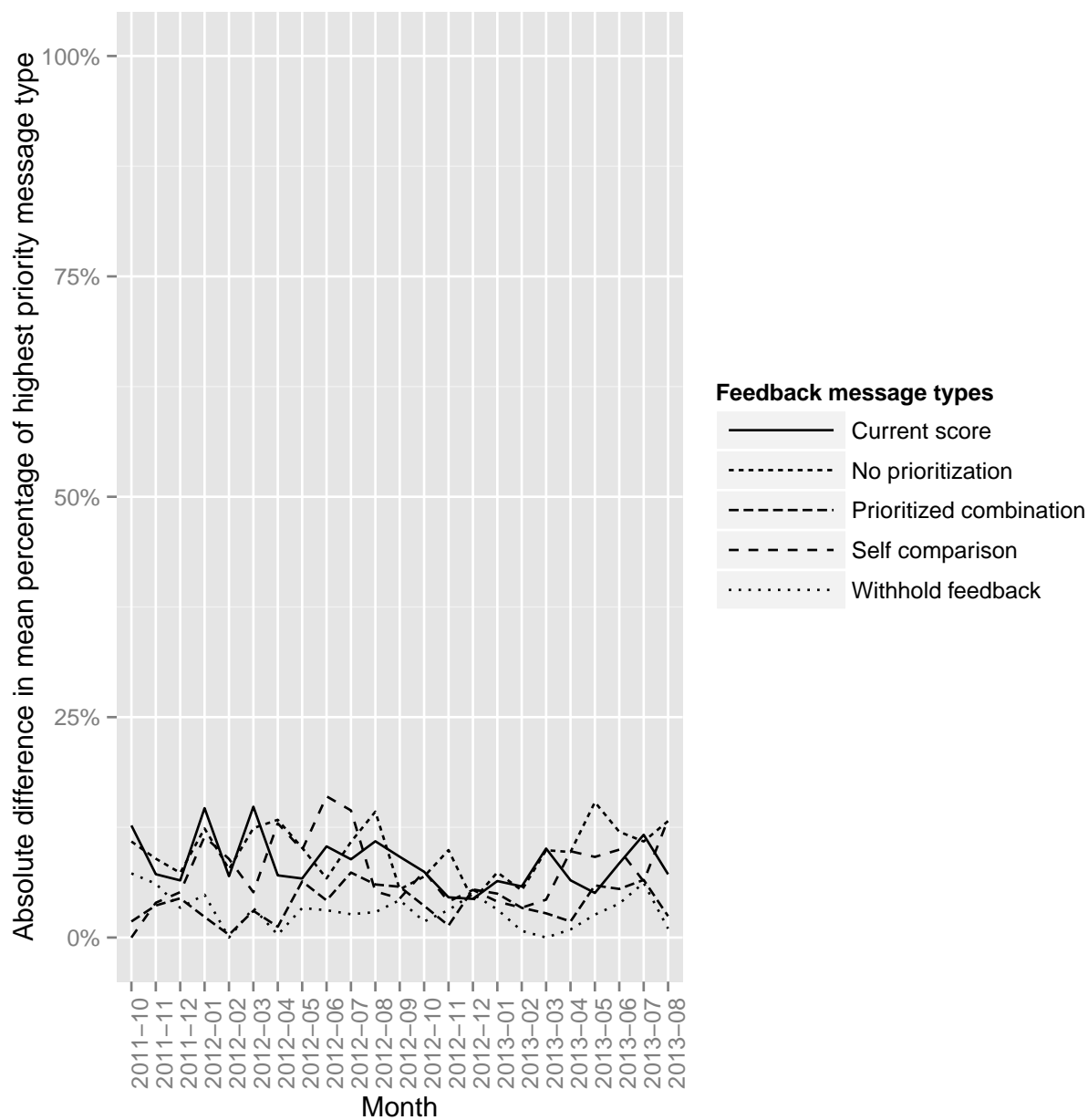


Figure 30: Mean absolute percentage change in messages for patient weight recording.

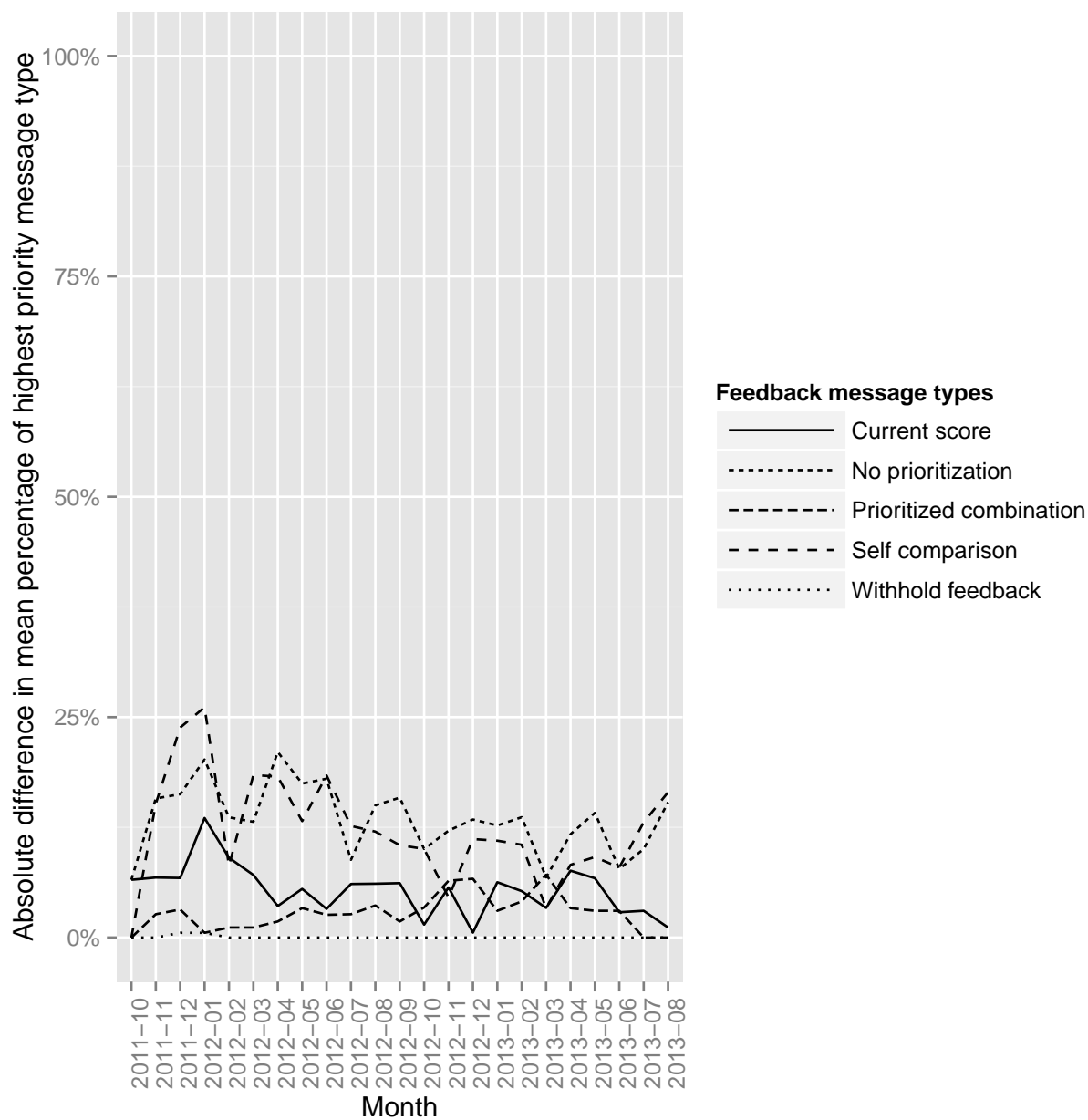


Figure 31: Mean absolute percentage change in messages for WHO clinical staging.

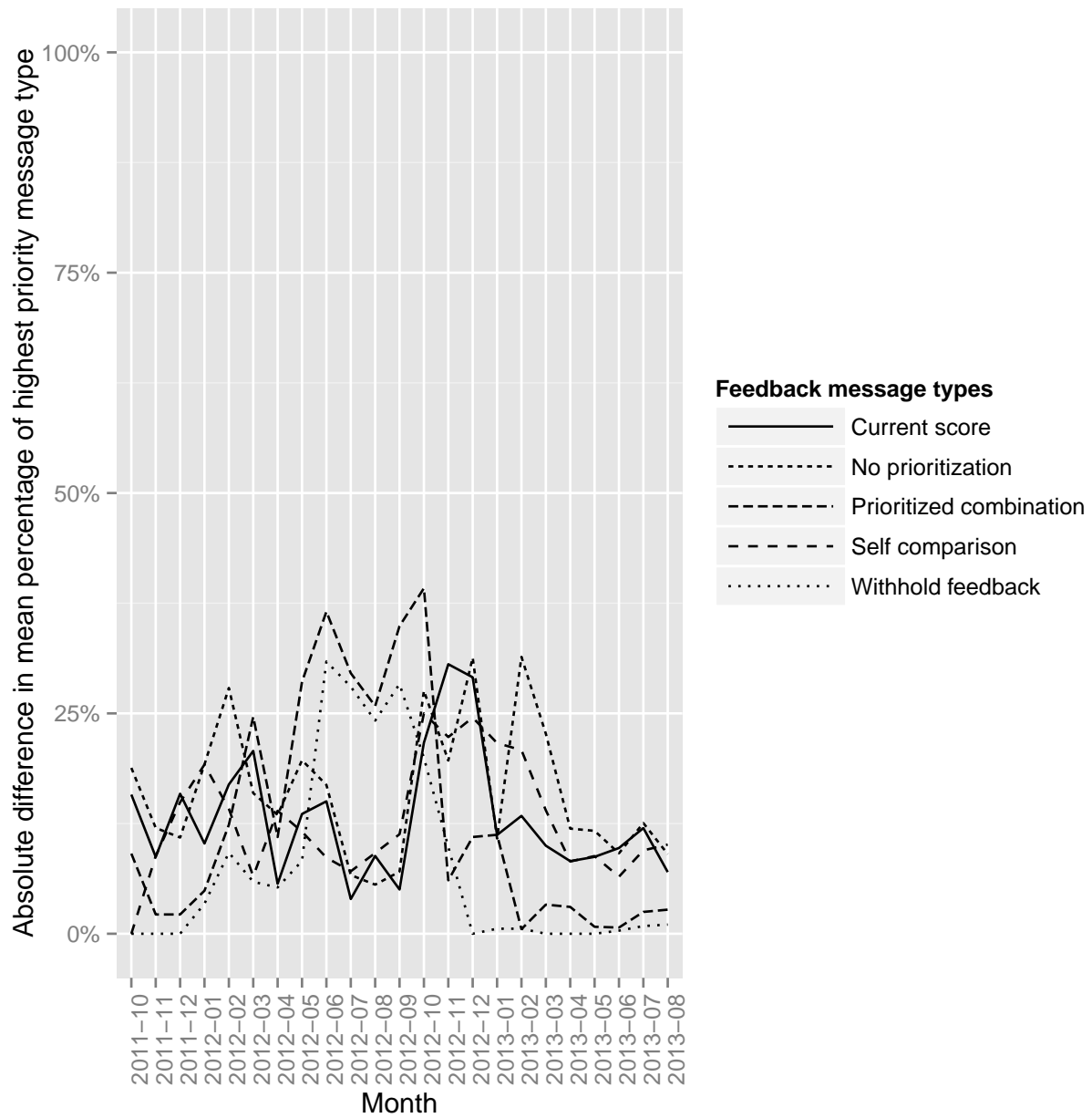


Figure 32: Mean absolute percentage change in messages for CPT prescribing.

14.3 DISCUSSION

The results of this study answer several important questions about the use of EMR data to generate tailored performance feedback messages in a low-resource setting. Most significantly, I identified an opportunity to use existing EMR data to routinely monitor individual clinical performance and provide credible, tailored feedback across a range of stable, guideline-based performance indicators in a low-resource setting. This system could be expected to generate individualized monthly reports for ART providers working at each site, with 35% of reports being tailored to optimize the effect of feedback on performance.

Although performance appears to have limited room for improvement in some ART clinics, I found that, in several clinics there are regular opportunities to provide individualized feedback to address performance gaps and possible data quality problems. These findings are significant because the existing National EMR infrastructure in Malawi would allow these reports to be generated automatically in every ART clinic using the EMR, requiring minimal additional resources. Moreover, such a system could generate feedback more rapidly than the current quarterly reporting schedule of the National ART monitoring and evaluation program.

I sought to understand if tailoring feedback messages can impact clinical performance by exploring differences in features of performance data that could be used for message tailoring. I found that there appear to be differences in the features of clinical performance data in Malawi that hold meaningful implications for the design of feedback messages. On average, based on a preliminary set of causal mechanisms offered by behavioral and cognitive theories, more than 50% of feedback messages for pediatric height recording could be tailored for individual or situational differences in performance. Similarly, close to an average of 50% of feedback messages could be tailored for differences in performance with regard to CPT prescribing. Where performance is higher, there appear to be fewer opportunities to tailor feedback messages. However, even the indicators having higher performance allowed for routine tailoring for approximately 25% of messages for weight recording on average, and for an average of 16.3% of messages for WHO clinical staging.

Another important finding was that the average percentage of messages routinely differed

across categories of message types by month by a moderate amount (Table 14). This finding suggests that message tailoring could be beneficial prior to the delivery of each monthly feedback message, rather than after a single initial assessment. This appears to be especially true for indicators with lower overall performance. For example, I found that, on average, the percentage of messages tailored for different categories changes by an average of 13% between months for pediatric height recording.

These findings are significant because they represent the first evaluation, to my knowledge, of a performance feedback tailoring knowledge-base that applies psychological theory for the purpose of automated feedback message tailoring. This approach represents a novel contribution that holds implications for related research in biomedical informatics, implementation science, and global health. In the field of biomedical informatics, this work contributes a novel class of knowledge-based system to support evidence-based care and quality improvement. A key implication for implementation science research is that automated tailoring systems could enable researchers to produce generalizable knowledge about how and when feedback interventions are effective across diverse clinical settings, to better understand the effectiveness of such interventions. In the domain of global health, this work represents the first supervision tool of its kind to support supervision shortages and mitigate high staff turnover. The significance of these findings increase with the increasing use of eHealth and resulting availability of eHealth data that can be used to generate performance feedback.

An important issue that requires further study is the use of automated feedback message tailoring in conjunction with eHealth data quality analysis. The performance indicator data we used can represent either actual performance of individuals or problems with the recording of clinical data that give the appearance of performance problems. In either case the use of performance measurement and provision of individual-level feedback may be used to work towards improved quality of care provided to patients.

14.4 LIMITATIONS

This research has several limitations. Firstly, I identified opportunities to provide feedback where performance is low, but I did not evaluate the degree to which low clinical performance is caused by environmental factors that individualized, tailored feedback cannot directly impact. For example, low performance of CPT prescribing is likely to be largely attributable to a shortage of CPT drugs, and tailored feedback can not directly influence performance in the absence of necessary material resources. However, I believe that the tailoring of feedback messages in these cases could nevertheless have positive impacts on the quality of the clinical data and indirectly on clinical performance. For example, I witnessed a national shortage of CPT drug contributing to data quality problems while observing the use of the EMR in the study described in Chapter 12. In this case, showing performance gaps between peers could highlight the variable practices in the use of the EMR that contributes to data quality problems for multiple purposes of data use. If performance feedback can be used to improve data quality, this could in turn increase the ability of supervisors to understand the impact of resource shortages on clinical performance and perhaps improve the allocation of resources.

Another limitation of this analysis is that the application of theory within tailoring rules was not rigorously validated. I applied theory using my knowledge gained from a review of the literature and in consultation with a cognitive psychologist who has expertise in the application of psychological theory to the design of clinical audit and feedback.

Finally, the classification thresholds that I used for this analysis, listed in section 13.2.3 were chosen based on my understanding of the clinical environment and my perceptions of the expectations of system stakeholders, rather than empirical research. For example, I classified low group performance as an average performance below 50%, but it is likely that the actual thresholds for low group performance may vary across ART clinics, and across performance indicators. In the case of WHO clinical staging, a threshold for low group performance might be set much higher for most clinics because there are no valid exceptions to the guideline recommendation for this behavior. In the case of CPT prescribing there are valid exceptions, for patients with allergies to CPT drug ingredients, therefore a lower threshold would be expected. To address this limitation, I chose classification thresholds

that err on the side of a lower bound, meaning that actual expectations are likely to lead to greater variability of tailored messages than the variability I found in this study.

14.5 CONCLUSION

Routine, individually-tailored performance feedback can be generated using existing EMR data at a national level in ART clinics in Malawi. I found that feedback reports could be routinely generated for ART providers in all clinics, with reports identifying approximately one performance gap of 10% or greater between peers at each site, per performance indicator, per month. Furthermore, I found that, using the prototype feedback tailoring knowledge-base, 35% of reports could be tailored to improve the effect of feedback on performance. There appear to be routine and promising opportunities for clinical supervisors to use a knowledge-based feedback tailoring tool to improve the effect of feedback on clinical performance. Future research should study the use of a prototype feedback tailoring system and its impact on the delivery of feedback messages and clinical performance.

15.0 DISCUSSION

This research has explored the potential for eHealth data to be used to generate automated performance feedback in clinical settings. In conducting these studies, I have established a foundation of knowledge for the development of automated performance feedback tools that could be used to improve clinical performance. Most significantly, this research contributes a novel approach for delivering clinical performance feedback: automated tailoring of feedback messages.

I began by seeking to generate meaningful automated performance feedback for health-care providers in Malawi using existing EMR data and a clinical guideline document. To accomplish this goal, I developed a method for identifying EMR-auditable, guideline-based performance indicators, and found 21 such indicators for use in ART clinics in Malawi (Chapter 11).¹⁸⁵ Perhaps the most significant contribution of this study was its demonstration of a method for opportunistically identifying performance measures that could be used to generate individualized feedback using EMR data in a low-resource setting. This study did not consider the complexity of the clinical environment and differences across individuals who receive feedback, but it prepared important ground work to establish the potential for opportunities to provide automated performance feedback.

The subsequent study reported in Chapter 12 explored the complexity created by the environment and by individual and situational differences in barriers to behavior change. To understand how to successfully design and implement a software application that can generate automated feedback using the performance indicators I identified in Chapter 11, I qualitatively evaluated barriers to using EMR data to provide meaningful performance feedback in ART clinics in Malawi. I identified four factors as key barriers to the implementation of automated audit and feedback: provider rotations, disruptions to care processes, eHealth

user acceptance, and performance indicator lifespan. This study contributed a novel scientific model for representing the complex factors relevant to implementing performance feedback in low-resource clinical settings. By placing the previously identified performance measures in context, this research helped to ground the design work for an automated feedback system in the uncertain and complex reality of the clinical setting.

I used the barriers we identified to inform the design of a novel, knowledge-based system that may mitigate the complexity of clinical settings to provide meaningful automated feedback: an automated feedback message tailoring system. This system could be used by clinical supervisors to select feedback messages from a menu of messages that are tailored for the individual and the environment (Chapter 14). To create a feedback tailoring knowledge-base, I identified theoretical constructs explaining the influence of individual and situational differences on the effectiveness of performance feedback, and developed a mechanism for the automated tailoring of feedback messages to accommodate these differences prior to the delivery of each message.

Finally, to understand the potential impact of this novel system, I retrospectively analyzed two years of de-identified EMR data from ART clinics in Malawi. Using the approach established in Chapter 11, based on my understanding of the setting established in Chapter 12, I formatively evaluated the system design (Chapter 14) by calculating how frequently and how variably tailored feedback messages could have been generated. I found that performance gaps could be routinely identified at a rate of approximately one per month, per performance indicator, per site, and approximately 35% of reports could have a menu of tailored messages automatically tailored to improve the effect of feedback on performance.

These findings, while exploratory and preliminary, are timely because of the growing availability of eHealth data for analysis in clinical settings, and the recognition of building the learning health system as a key challenge for improving the quality of healthcare.⁴ I envision an automated feedback tailoring system as a tool support that can support a culture of continuous learning, and to have increased impact during implementation of new knowledge in clinical settings.

15.1 LIMITATIONS

This research has important limitations. The primary limitation for this work is that the foundational elements of the system design, while grounded in significant research effort, have not yet been evaluated in a laboratory or field environment. A critical next step for this research is the evaluation of a prototype feedback message tailoring system with clinical supervisors and ART providers. A second limitation is that the factors we identified as barriers to the implementation of such a system may not be comprehensive, therefore we can not know what further barriers to system implementation exist until we successfully implement a prototype system. Third, the approach we have selected for the use of theory is novel in itself, and while it represents an innovative and promising direction for the systematic use of theory to improve the effectiveness of implementation science research, it does not represent a time-tested and widely accepted method for using psychological theory. Fourth, our work is limited to ART clinics in Malawi, therefore our findings can not be generalized to other low-resource settings and clinical settings in high-income countries. Evaluating this approach in different contexts is necessary to overcome this limitation. Finally, the costs of developing and maintaining a knowledge-based feedback tailoring system are unclear, therefore it is not certain that the creation of this type of system will add value to the existing eHealth infrastructure within a clinic.

15.2 FUTURE WORK

This research reveals many potential areas of inquiry that future work should explore, including the use of theory, domains in which this kind of tailoring system can be applied, and the costs of building and maintaining such systems. Regarding using theory to guide feedback message tailoring, we do not know the extent to which additional theoretical constructs and causal mechanisms are relevant to inform feedback tailoring. The representation of expert knowledge about cognitive and behavioral theories having implications for automated feedback message tailoring is a largely unexplored domain.

Perhaps the most promising direction for this research is to complement this work with a data-driven approach that informs the use of theory, rather than focusing on theory-driven evaluations of feedback tailoring alone. The degree to which machine learning approaches could guide tailoring algorithms once a message tailoring system was routinely used is another area of inquiry that holds significant potential. By putting the tool I have proposed to build into the hands of clinical supervisors and observing which messages they choose and how effective the use of each kind of message is, we may be able to more rapidly optimize feedback message tailoring in a more relevant way to the context of each specific intervention. A particularly exciting goal for an automated feedback tailoring is to create systems that can intelligently optimize feedback recommendations by learning from experience. I believe is feasible accomplish this goal using existing eHealth and informatics methods.

An example of an approach to developing a system that could intelligently optimize feedback tailoring recommendations would be to develop a probabilistic network that models the effect of specific feedback messages on performance for a given individual and situation. The development of such a network could use a data-driven Bayesian structure-learning approach, informed by theoretical causal mechanisms that are believed to significantly influence performance in the clinical setting in which the model is being evaluated. COM-B could provide organizing frameworks for inferring the probability of individual barriers to behavior change. For example, a task might require very little new knowledge or skill to perform successfully, therefore a prior probability of a capability-associated barrier would be low for all individuals performing the task. However, if an individual consistently did not perform the task without indication of opportunity or motivational barriers to performance, the probability of a capability barrier would be increased. This would cause the message tailoring system to prioritize messages tailored for an individual believed to have a capability-associated barrier to improving performance. These probabilities could potentially be learned from the data if supervisors were to use a tailoring menu to guide the provision of individualized performance feedback, indicating their beliefs about the presence of capability, opportunity, and motivational barriers when selecting each specific message.

The domains in which a feedback tailoring system might be useful are many. Within healthcare, performance feedback is provided across nearly every medical specialty for qual-

ity improvement and the implementation of evidence-based practice, in high-, low-, and middle-income countries. Beyond the use of performance feedback for healthcare professional behavior change, such a system may be particularly relevant for medical education settings, where learning-focused feedback is routinely given. For healthcare providers who give persuasive feedback to patients based on their lifestyle, such a system could potentially facilitate improved patient-provider communication. Beyond the domain of healthcare, performance feedback tailoring systems could be relevant for many types of organizations where routine performance feedback is provided, such as organizations in education, industry, and government.

16.0 CONCLUSIONS

This work has been motivated by gaps between medical evidence and the decisions made by healthcare professionals, by the growing availability of eHealth data, and by the mystery surrounding the effect of feedback interventions on performance. This research has contributed novel insights into the development of tools that show promise for overcoming the challenges associated with each of these motivating factors within the context of low-resource settings, where healthcare provider supervision and performance are suboptimal.

In this research, I first established the feasibility of measuring individual performance using EMR data in a low-resource settings by deriving performance indicators from a simplified clinical practice guideline and identifying representative EMR data to measure performance. Next I identified and described key contextual barriers that must be overcome to generate automated AF in public hospitals in Malawi. To accommodate these barriers, I designed a novel, theory-informed feedback message tailoring system that may enable clinical supervisors to routinely tailor feedback. Finally, I identified existing opportunities to routinely use such a tailoring system for healthcare providers in ART clinics in Malawi. This work has created a new foundation of knowledge for the development of a feedback message tailoring tool for improving the quality of care. I believe that this work holds high potential for broadly impacting clinical learning and behavior change to enable patients to receive the best care possible.

APPENDIX

INTERVIEW GUIDE

The interview guide included below is described in Chapter 12.

Factors affecting implementation of an automated audit and feedback system

Interview Guide: Health worker

Introduction

Before we begin I want to let you know that we are interested in your comfort in this interview and we are here to talk about anything you would like to talk about. Do I have your permission to turn the recorder on now?

Construct/Concepts	Questions	Probes
Warmup / Rapport Building	1. Could you begin by telling me a little about the work you do and your role in the ART clinic?	<ul style="list-style-type: none"> What areas do you work in other than the ART clinic? How many ART clinic days are there at your facility?
Sources of feedback	2. What are the ways you can know if you are following the ART guidelines?	<ul style="list-style-type: none"> After you go to ART training, when you come back to your clinic, how do you know that you are applying what you have learned? What differences are there between feedback received during quarterly supervision compared to other supervision visits? (If electronic records are mentioned) Is there a way that the electronic system is helping you to know how you are following the guidelines, or is that not part of the use of the electronic system?
Source credibility	3. How well can you trust the information in the EDS?	<ul style="list-style-type: none"> What would you say about how well you can rely on the information in the EDS? (and other sources if they are mentioned)
Feedback quality	4. Can you tell me a story about a time that you received ART guideline adherence feedback? 5. What opportunities do you see to improve the ART guideline adherence feedback you receive from supervisors?	<ul style="list-style-type: none"> What would you say are the strengths of the ART guideline adherence feedback you receive from supervisors? How frequently do you receive helpful feedback?

Factors affecting implementation of an automated audit and feedback system

Construct/Concepts	Questions	Probes
Feedback features	<p>6. How much feedback do you receive about your individual performance?</p> <p>7. Do you usually receive feedback verbally, in writing, or other ways?</p>	<ul style="list-style-type: none"> Are there differences in the manner that different supervisors provide feedback? What are the tasks that you do not receive enough feedback about?
Perceived accuracy/Acceptance	<p>8. What do you personally do with the feedback that you receive from supervisors?</p>	<ul style="list-style-type: none"> What would you say about how accurate performance feedback from supervisors is? What would you say about how you are accepting of the feedback you receive?
Relative priority of performance feedback / Tension for change	<p>9. How important is it to you to respond to the feedback you receive from supervisors?</p>	<ul style="list-style-type: none"> (If indication that responding to feedback is not important) How important is it to you to follow the ART guidelines?
<p>Motivation:</p> <ul style="list-style-type: none"> - Desire to respond - Intent to respond - Regulatory Focus - Feedback self-efficacy 	<p>10. Why is it important to you to respond to the feedback you receive from supervisors?</p> <p>11. What might prevent you from making changes based on the feedback you receive?</p> <p>12. What are the consequences of ignoring feedback that from a supervisor?</p>	<ul style="list-style-type: none"> Can you think of a time when you understood how to respond to feedback, but you weren't able to?
Wrap-up	<p>13. If you could use the EDS to get a feedback report about your performance according to the ART guidelines, what would you want to see?</p> <p>14. Is there anything else I need to know to understand your thoughts and feelings about the performance feedback that you receive?</p>	<ul style="list-style-type: none"> What information would you be most interested to receive about your performance as a (clinician/nurse)?

BIBLIOGRAPHY

- [1] Davis D, Evans M, Jadad A, Perrier L, Rath D, Ryan D, et al. The case for knowledge translation: shortening the journey from evidence to effect. *BMJ (Clinical research ed)*. 2003 Jul;327(7405):33–35.
- [2] McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. *The New England Journal of Medicine*. 2003 Jun;348(26):2635–2645.
- [3] Rowe AK, de Savigny D, Lanata CF, Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *The Lancet*. 2005 Sep;366(9490):1026–1035. Available from: <http://www.sciencedirect.com/science/article/B6T1B-4GV8TBC-1/2/0607c824b3ec3b90540fe31c646160b6>.
- [4] Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, D.C.: The National Academies Press: IOM (Institute of Medicine); 2012.
- [5] Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The Impact of eHealth on the Quality and Safety of Health Care: A Systematic Overview. *PLoS Med*. 2011 Jan;8(1):e1000387. Available from: <http://dx.doi.org/10.1371/journal.pmed.1000387>.
- [6] Wright A, Henkin S, Feblowitz J, McCoy AB, Bates DW, Sittig DF. Early Results of the Meaningful Use Program for Electronic Health Records. *New England Journal of Medicine*. 2013;368(8):779–780. Available from: <http://www.nejm.org/doi/full/10.1056/NEJMc1213481>.
- [7] Blaya JA, Fraser HSF, Holt B. E-health technologies show promise in developing countries. *Health Affairs (Project Hope)*. 2010 Feb;29(2):244–251. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20348068>.
- [8] Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Affairs (Project Hope)*. 2014 Jul;33(7):1163–1170.
- [9] Brehaut JC, Eva KW. Building theories of knowledge translation interventions: Use

- the entire menu of constructs. *Implementation Science*. 2012 Nov;7(1):114. Available from: <http://www.implementationscience.com/content/7/1/114/abstract>.
- [10] Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane database of systematic reviews* (Online). 2012;6:CD000259.
 - [11] Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implementation Science*. 2014 Jan;9(1):14. Available from: <http://www.implementationscience.com/content/9/1/14/abstract>.
 - [12] Colquhoun HL, Brehaut JC, Sales A, Ivers N, Grimshaw J, Michie S, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implementation Science*. 2013 Jun;8(1):66. Available from: <http://www.implementationscience.com/content/8/1/66/abstract>.
 - [13] Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical care*. 2013 Aug;51(8 Suppl 3):S30–37.
 - [14] Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Medical care*. 2012 Jul;50 Suppl:S21–29.
 - [15] Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. *J Manage Inf Syst*. 1996 Mar;12(4):5–33. Available from: <http://dl.acm.org/citation.cfm?id=1189570.1189572>.
 - [16] Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. *The Medical journal of Australia*. 2004 Mar;180(6 Suppl):S57–60.
 - [17] Fretheim A, Oxman AD, Flottorp S. Improving prescribing of antihypertensive and cholesterol-lowering drugs: a method for identifying and addressing barriers to change. *BMC Health Services Research*. 2004 Sep;4(1):23. Available from: <http://www.biomedcentral.com/1472-6963/4/23/abstract>.
 - [18] Grol R, Wensing M, Eccles M, Davis D. *Improving Patient Care: The Implementation of Change in Health Care*. John Wiley & Sons; 2013.
 - [19] Plsek PE, Greenhalgh T. The challenge of complexity in health care. *BMJ : British Medical Journal*. 2001 Sep;323(7313):625–628. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1121189/>.
 - [20] Weiner MG, Embi PJ. Toward Reuse of Clinical Data for Research and Quality Improvement: The End of the Beginning? *Annals of Internal*

- Medicine. 2009 Sep;151(5):359–360. Available from: <http://dx.doi.org/10.7326/0003-4819-151-5-200909010-00141>.
- [21] Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. *Journal of medical Internet research*. 2005;7(1):e1.
 - [22] Seebregts CJ, Mamlin BW, Biondich PG, Fraser HSF, Wolfe BA, Jazayeri D, et al. Human Factors for Capacity Building. Lessons learned from the OpenMRS Implementers Network. *Yearbook of Medical Informatics*. 2010;p. 13–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20938564>.
 - [23] Tierney WM, Achieng M, Baker E, Bell A, Biondich P, Braitstein P, et al. Experience implementing electronic health records in three East African countries. *Studies in Health Technology and Informatics*. 2010;160(Pt 1):371–375. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20841711>.
 - [24] Kanter AS, Negin J, Olayo B, Bukachi F, Johnson E, Sachs SE. Millennium Global Village-Net: bringing together Millennium Villages throughout sub-Saharan Africa. *International Journal of Medical Informatics*. 2009 Dec;78(12):802–807. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19766532>.
 - [25] Mohammed-Rajput NA, Smith DC, Mamlin B, Biondich P, Doebbeling BN, Open MRS Collaborative Investigators. OpenMRS, a global medical records system collaborative: factors influencing successful implementation. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium*. 2011;2011:960–968.
 - [26] Douglas GP, Gadabu OJ, Joukes S, Mumba S, McKay MV, Ben-Smith A, et al. Using Touchscreen Electronic Medical Record Systems to Support and Monitor National Scale-Up of Antiretroviral Therapy in Malawi. *PLoS Med*. 2010;7(8):e1000319. Available from: <http://dx.doi.org/10.1371/journal.pmed.1000319>.
 - [27] Forster M, Bailey C, Brinkhof MWG, Graber C, Boulle A, Spohr M, et al. Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings. *Bulletin of the World Health Organization*. 2008 Dec;86(12):939–947. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19142294>.
 - [28] AbouZahr C, Boerma T. Health information systems: the foundations of public health. *Bulletin of the World Health Organization*. 2005 Aug;83(8):578–583. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16184276>.
 - [29] Braa J, Kanter AS, Lesh N, Crichton R, Jolliffe B, SÃebÃy J, et al. Comprehensive yet scalable health information systems for low resource settings: a collaborative effort in sierra leone. *AMIA Annual Symposium Proceedings / AMIA Symposium AMIA Symposium*. 2010;2010:372–376. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21347003>.

- [30] Powell A, Davies H, Thomson R. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Quality & safety in health care*. 2003 Apr;12(2):122–128. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1743685/>.
- [31] Maydanchik A. Data quality assessment. Bradley Beach, N.J.: Technics Publications; 2007. Available from: <http://www.books24x7.com/marc.asp?bookid=34410>.
- [32] Wand Y, Wang RY. Anchoring Data Quality Dimensions in Ontological Foundations. *Commun ACM*. 1996 Nov;39(11):86–95. Available from: <http://doi.acm.org/10.1145/240455.240479>.
- [33] The World Health Report 2006: Working Together For Health. WHO; 2006. Available from: <http://www.who.int/whr/2006/en/>.
- [34] The Global Burden of Disease: 2004 Update. WHO; 2008. Available from: http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html.
- [35] Malawi, UNAIDS Country profile; 2012. Available from: <http://www.unaids.org/en/regionscountries/countries/malawi/>.
- [36] USAID HIV/AIDS Health Profile for Malawi - September 2010. USAID; 2010. Available from: http://www.usaid.gov/our_work/global_health/aids/Countries/africa/malawi.html.
- [37] Integrated HIV Program Report, January - March 2014. Government of Malawi, Ministry of Health; 2014.
- [38] Harries AD, Zachariah R, Jahn A, Schouten EJ, Kamoto K. Scaling up antiretroviral therapy in Malawi-implications for managing other chronic diseases in resource-limited countries. *Journal of Acquired Immune Deficiency Syndromes (1999)*. 2009 Nov;52 Suppl 1:S14–16. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19858929>.
- [39] Atun R, Bataringaya J. Building a durable response to HIV/AIDS: implications for health systems. *Journal of Acquired Immune Deficiency Syndromes (1999)*. 2011 Aug;57 Suppl 2:S91–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21857305>.
- [40] Harries AD, Gomani P, Teck R, de Teck OA, Bakali E, Zachariah R, et al. Monitoring the response to antiretroviral therapy in resource-poor settings: the Malawi model. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 2004 Dec;98(12):695–701.
- [41] Libamba E, Makombe S, Mhango E, de Ascurra Teck O, Limbambala E, Schouten EJ, et al. Supervision, monitoring and evaluation of nationwide scale-up of antiretroviral

- therapy in Malawi. *Bulletin of the World Health Organization*. 2006 Apr;84(4):320–326. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16628306>.
- [42] Schouten EJ, Jahn A, Ben-Smith A, Makombe SD, Harries AD, Aboagye-Nyame F, et al. Antiretroviral drug supply challenges in the era of scaling up ART in Malawi. *Journal of the International AIDS Society*. 2011;14 Suppl 1:S4.
- [43] Rabin BA, Brownson RC, Haire-Joshu D, Kreuter MW, Weaver NL. A glossary for dissemination and implementation research in health. *Journal of Public Health Management and Practice: JPHMP*. 2008 Apr;14(2):117–123. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18287916>.
- [44] McKibbin KA, Lokker C, Wilczynski NL, Ciliska D, Dobbins M, Davis DA, et al. A cross-sectional study of the number and frequency of terms used to refer to knowledge translation in a body of health literature in 2006: a Tower of Babel? *Implementation science: IS*. 2010;5:16.
- [45] Straus SE, Tetroe J, Graham ID. Introduction Knowledge translation: What it is and what it isn't. In: MSc SESM FRCPC, Tetroe J, FCAHS IDG, editors. *Knowledge Translation in Health Care*. John Wiley & Sons, Ltd; 2013. p. 1–13. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/9781118413555.ch01/summary>.
- [46] Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? *The Journal of Continuing Education in the Health Professions*. 2006;26(1):13–24.
- [47] Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science: IS*. 2009;4:50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19664226>.
- [48] English M, Nzinga J, Mbindyo P, Ayieko P, Irimu G, Mbaabu L. Explaining the effects of a multifaceted intervention to improve inpatient care in rural Kenyan hospitals—interpretation based on retrospective examination of data from participant observation, quantitative and qualitative studies. *Implementation Science: IS*. 2011;6:124. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22132875>.
- [49] Padian NS, Holmes CB, McCoy SI, Lyerla R, Bouey PD, Goosby EP. Implementation science for the US President's Emergency Plan for AIDS Relief (PEPFAR). *Journal of Acquired Immune Deficiency Syndromes (1999)*. 2011 Mar;56(3):199–203. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21239991>.
- [50] Harries AD, Makombe SD, Schouten EJ, Jahn A, Libamba E, Kamoto K, et al. How operational research influenced the scale up of antiretroviral therapy in Malawi. *Health Care Management Science*. 2011 Nov; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22113539>.

- [51] Lowrance DW, Makombe S, Harries AD, Shiraishi RW, Hochgesang M, Aberle-Grasse J, et al. A public health approach to rapid scale-up of antiretroviral treatment in Malawi during 2004-2006. *Journal of Acquired Immune Deficiency Syndromes* (1999). 2008 Nov;49(3):287–293. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18845953>.
- [52] Zachariah R, Ford N, Philips M, Lynch S, Massaquoi M, Janssens V, et al. Task shifting in HIV/AIDS: opportunities, challenges and proposed actions for sub-Saharan Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 2009 Jun;103(6):549–558. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18992905>.
- [53] Mullan F, Frehywot S. Non-physician clinicians in 47 sub-Saharan African countries. *Lancet*. 2007 Dec;370(9605):2158–2163. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17574662>.
- [54] Morris MB, Chapula BT, Chi BH, Mwango A, Chi HF, Mwanza J, et al. Use of task-shifting to rapidly scale-up HIV treatment services: experiences from Lusaka, Zambia. *BMC Health Services Research*. 2009;9:5. PMC2628658. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2628658>.
- [55] Brennan A, Long L, Maskew M, Sanne I, Jaffray I, Macphail P, et al. Outcomes of stable HIV-positive patients down-referred from doctor-managed ART clinics to nurse-managed primary health clinics for monitoring and treatment. *AIDS (London, England)*. 2011 Aug; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21832934>.
- [56] Callaghan M, Ford N, Schneider H. A systematic review of task- shifting for HIV treatment and care in Africa. *Human Resources for Health*. 2010;8:8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20356363>.
- [57] Task shifting: rational redistribution of tasks among health workforce teams. Global recommendations and guidelines. Geneva, Switzerland: World Health Organization [WHO], Health Systems and Services; 2008. Available from: <http://www.who.int/entity/healthsystems/TTR-TaskShifting.pdf>.
- [58] Rowe AK, Onikpo F, Lama M, Deming MS. The rise and fall of supervision in a project designed to strengthen supervision of Integrated Management of Childhood Illness in Benin. *Health Policy and Planning*. 2010 Mar;25(2):125–134. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19923206>.
- [59] Bosch-Capblanch X, Liaqat S, Garner P. Managerial supervision to improve primary health care in low- and middle-income countries. *Cochrane Database of Systematic Reviews* (Online). 2011;9:CD006413. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21901704>.
- [60] Frimpong JA, Helleringer S, Awoonor-Williams JK, Yeji F, Phillips JF. Does supervision improve health worker productivity? Evidence from the Upper East Region of

- Ghana. Tropical Medicine & International Health: TM & IH. 2011 Jul; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21729221>.
- [61] Marquez L, Kean L. MAQ Papers: Making Supervision Supportive and Sustainable: New Approaches to Old Problems. Management Sciences for Health [MSH]; 2002.
 - [62] WHO recommendations for clinical mentoring to support scale-up of HIV care, antiretroviral therapy and prevention in resource-constrained settings. World Health Organization [WHO]; 2006. Available from: <http://www.who.int/hiv/pub/meetingreports/clinicalmentoring/en/>.
 - [63] Dieleman M, Gerretsen B, van der Wilt GJ. Human resource management interventions to improve health workers' performance in low and middle income countries: a realist review. Health Research Policy and Systems / BioMed Central. 2009;7:7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19374734>.
 - [64] Quarterly HIV Programme Report 2011 Q2. Malawi Ministry of Health, Department of HIV and AIDS; 2011. Available from: <http://www.hivunitmohmw.org/Main/AntiretroviralTherapy>.
 - [65] I-TECH. Clinical Mentoring Toolkit; 2008. Available from: <http://www.go2itech.org/HTML/CM08/index.html>.
 - [66] Norman G. Building on experience—the development of clinical reasoning. The New England Journal of Medicine. 2006 Nov;355(21):2251–2252. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17124025>.
 - [67] Chacko KM, Aagard E, Irby D. Teaching models for outpatient medicine. The Clinical Teacher. 2007 Jun;4(2):82–86. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1743-498X.2007.00153.x/abstract>.
 - [68] Field MJ, Lohr KN, editors. Guidelines for Clinical Practice: From Development to Use. Washington, D.C.: The National Academies Press; 1992.
 - [69] Margolis CZ, Cretin S. Implementing clinical practice guidelines. AHA Press; 1999.
 - [70] Shapiro DW, Lasker RD, Bindman AB, Lee PR. Containing costs while improving quality of care: the role of profiling and practice guidelines. Annual Review of Public Health. 1993;14:219–241. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8323588>.
 - [71] Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ (Clinical Research Ed). 1999 Feb;318(7182):527–530. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10024268>.

- [72] Norris SL, Holmer HK, Ogden LA, Burda BU. Conflict of interest in clinical practice guideline development: a systematic review. *PloS One*. 2011;6(10):e25153. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22039406>.
- [73] SIGN 50: A guideline developer's handbook; 2011. Available from: <http://www.sign.ac.uk/guidelines/fulltext/50/index.html>.
- [74] Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality & Safety in Health Care*. 2003 Feb;12(1):18–23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12571340>.
- [75] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical Research Ed)*. 2008 Apr;336(7650):924–926. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18436948>.
- [76] Shiffman RN, Dixon J, Brandt C, Essaihi A, Hsiao A, Michel G, et al. The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. *BMC Medical Informatics and Decision Making*. 2005;5:23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16048653>.
- [77] Wabitsch R, Margolis CZ, Malkin JE, Abu-Saad K, Waksman J. Evaluating the process of tailoring the global HIV/AIDS practice guidelines for use in developing countries. *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care / ISQua*. 1998 Apr;10(2):147–154. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9690888>.
- [78] English M, Opiyo N. Getting to grips with GRADE-perspective from a low-income setting. *Journal of Clinical Epidemiology*. 2011 Jul;64(7):708–710. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21316192>.
- [79] Fretheim A, SchÄijnemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 15. Disseminating and implementing guidelines. *Health Research Policy and Systems / BioMed Central*. 2006;4:27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17156459>.
- [80] Oxman AD, Lavis JN, Fretheim A. Use of evidence in WHO recommendations. *Lancet*. 2007 Jun;369(9576):1883–1889. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17493676>.
- [81] Wiysonge CS, Hussey GD. Use of systematic reviews in WHO recommendations. *Lancet*. 2011 Jun;377(9782):2006. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21665038>.
- [82] Johansson KA, Jerene D, Norheim OF. National HIV treatment guidelines in Tanzania and Ethiopia: are they legitimate rationing tools? *Journal of Medical Ethics*.

- 2008 Jun;34(6):478–483. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18511624>.
- [83] De Weggheleire A, Bortolotti V, Zolfo M, Crowley S, Colebunders R, Riedner G, et al. Challenges in developing national HIV guidelines: experience from the eastern Mediterranean. *Bulletin of the World Health Organization*. 2011 Jun;89(6):442–450. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21673860>.
 - [84] English M, Scott JAG. What Is the Future for Global Case Management Guidelines for Common Childhood Diseases? *PLoS Medicine*. 2008 Dec;5(12):e241 EP –. Available from: <http://dx.doi.org/10.1371%2Fjournal.pmed.0050241>.
 - [85] Malawi Antiretroviral Therapy Treatment Programme: Quarterly Report: Results up to December 2010. Ministry of Health Malawi, Dept. of HIV and AIDS; 2011.
 - [86] Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants. World Health Organization [WHO]; 2010. Available from: <http://www.who.int/hiv/pub/mtct/antiretroviral2010/en/index.html>.
 - [87] Butzlaff M, Vollmar HC, Floer B, Konecny N, Isfort J, Lange S. Learning with computerized guidelines in general practice?: A randomized controlled trial. *Family Practice*. 2004 Apr;21(2):183–188. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15020389>.
 - [88] Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA: The Journal of the American Medical Association*. 1999 Oct;282(15):1458–65. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10535437>.
 - [89] Francke AL, Smit MC, de Veer AJE, Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. *BMC Medical Informatics and Decision Making*. 2008;8:38. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18789150>.
 - [90] Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *Journal of Clinical Epidemiology*. 1991;44(11):1271–1278. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1834807>.
 - [91] Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technology Assessment (Winchester, England)*. 2004 Feb;8(6):iii–iv, 1–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14960256>.
 - [92] Ayieko P, Ntoburi S, Wagai J, Opondo C, Opiyo N, Migiro S, et al. A multifaceted intervention to implement guidelines and improve admission paediatric care in Kenyan district hospitals: a cluster randomised trial. *PLoS Medicine*. 2011 Apr;8(4):e1001018. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21483712>.

- [93] Harries AD, Makombe SD, Libamba E, Schouten EJ. Why did the scale-up of HIV treatment work? A case example from Malawi. *Journal of Acquired Immune Deficiency Syndromes* (1999). 2011 Aug;57 Suppl 2:S64–67. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21857298>.
- [94] Rowe AK, Onikpo F, Lama M, Osterholt DM, Rowe SY, Deming MS. A multifaceted intervention to improve health worker adherence to integrated management of childhood illness guidelines in Benin. *American Journal of Public Health*. 2009 May;99(5):837–846. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19299681>.
- [95] Nzinga J, Mbindyo P, Mbaabu L, Warira A, English M. Documenting the experiences of health workers expected to implement guidelines during an intervention study in Kenyan hospitals. *Implementation Science*. 2009;4(1):44. Available from: <http://www.implementationscience.com/content/4/1/44>.
- [96] Nzinga J, Ntoburi S, Wagai J, Mbindyo P, Mbaabu L, Migiro S, et al. Implementation experience during an eighteen month intervention to improve paediatric and newborn care in Kenyan district hospitals. *Implementation Science: IS*. 2009;4:45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19627594>.
- [97] Walter ND, Lyimo T, Skarbinski J, Metta E, Kahigwa E, Flannery B, et al. Why first-level health workers fail to follow guidelines for managing severe disease in children in the Coast Region, the United Republic of Tanzania. *Bulletin of the World Health Organization*. 2009 Feb;87(2):99–107. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19274361>.
- [98] Barasa EW, Ayieko P, Cleary S, English M. A multifaceted intervention to improve the quality of care of children in district hospitals in Kenya: a cost-effectiveness analysis. *PLoS medicine*. 2012;9(6):e1001238.
- [99] Ivers NM, Tu K, Young J, Francis JJ, Barnsley J, Shah BR, et al. Feedback GAP: pragmatic, cluster-randomized trial of goal setting and action plans to increase the effectiveness of audit and feedback interventions in primary care. *Implementation science: IS*. 2013;8:142.
- [100] Chalouhi GE, Salomon LJ, Fontanges M, Althuser M, Haddad G, Scemama O, et al. Formative assessment based on an audit and feedback improves nuchal translucency ultrasound image quality. *Journal of ultrasound in medicine: official journal of the American Institute of Ultrasound in Medicine*. 2013 Sep;32(9):1601–1605.
- [101] Simunovic M, Coates A, Smith A, Thabane L, Goldsmith CH, Levine MN. Uptake of an innovation in surgery: observations from the cluster-randomized Quality Initiative in Rectal Cancer trial. *Canadian journal of surgery Journal canadien de chirurgie*. 2013 Dec;56(6):415–421.

- [102] Geller BM, Ichikawa L, Miglioretti DL, Eastman D. Web-based mammography audit feedback. *AJR American journal of roentgenology*. 2012 Jun;198(6):W562–567.
- [103] Foy R, Eccles MP, Jamtvedt G, Young J, Grimshaw JM, Baker R. What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review. *BMC Health Services Research*. 2005;5:50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16011811>.
- [104] Langley GJ, Moen R, Nolan KM, Nolan TW, Norman CL, Provost LP. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. 2nd ed. Jossey-Bass; 2009.
- [105] Donabedian A. *An Introduction to Quality Assurance in Health Care*. 1st ed. Oxford University Press, USA; 2002.
- [106] Weiss KB, Wagner R. Performance measurement through audit, feedback, and profiling as tools for improving clinical care. *Chest*. 2000 Aug;118(2 Suppl):53S–58S. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10940000>.
- [107] Hysong SJ. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Medical Care*. 2009 Mar;47(3):356–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19194332>.
- [108] Kluger AN, DeNisi A. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin* March 1996. 1996;119(2):254–284.
- [109] WHO. *WHO Global Strategy for Containment of Antimicrobial Resistance*. Geneva, Switzerland: World Health Organization; 2001. Available from: http://www.who.int/csr/resources/publications/drugresist/WHO_CDS_CSR_DRS_2001_2_EN/en/.
- [110] Goossens H, Ferech M, Vander Stichele R, Elseviers M, ESAC Project Group. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet*. 2005 Feb;365(9459):579–587.
- [111] Charani E, Edwards R, Sevdalis N, Alexandrou B, Sibley E, Mullett D, et al. Behavior change strategies to influence antimicrobial prescribing in acute care: a systematic review. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2011 Oct;53(7):651–662.
- [112] Davey P, Brown E, Charani E, Fenelon L, Gould IM, Holmes A, et al. Interventions to improve antibiotic prescribing practices for hospital inpatients. *The Cochrane database of systematic reviews*. 2013;4:CD003543.
- [113] CDC. *Core Elements of Hospital Antibiotic Stewardship Programs*. Atlanta, GA: US Department of Health and Human Services, CDC; 2014. Available from: <http://www.cdc.gov/getsmart/healthcare/implementation/core-elements.html>.

- [114] Arnold SR, Straus SE. Interventions to improve antibiotic prescribing practices in ambulatory care. The Cochrane database of systematic reviews. 2005;(4):CD003539.
- [115] Trautner BW, Kelly PA, Petersen N, Hysong S, Kell H, Liao KS, et al. A hospital-site controlled intervention using audit and feedback to implement guidelines concerning inappropriate treatment of catheter-associated asymptomatic bacteriuria. *Implementation science: IS*. 2011;6:41.
- [116] Nathwani D, Gray K, Borland H. Quality indicators for antibiotic control programmes. *The Journal of hospital infection*. 2002 Mar;50(3):165–169.
- [117] Baysari MT, Oliver K, Egan B, Li L, Richardson K, Sandaradura I, et al. Audit and feedback of antibiotic use: utilising electronic prescription data. *Applied clinical informatics*. 2013;4(4):583–595.
- [118] Wahlström R, Kounnavong S, Sisounthone B, Phanyanouvong A, Southammavong T, Eriksson B, et al. Effectiveness of feedback for improving case management of malaria, diarrhoea and pneumonia—a randomized controlled trial at provincial hospitals in Lao PDR. *Tropical Medicine & International Health: TM & IH*. 2003 Oct;8(10):901–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14516301>.
- [119] van den Akker T, Mwangomba B, Irlam J, van Roosmalen J. Using audits to reduce the incidence of uterine rupture in a Malawian district hospital. *International Journal of Gynaecology and Obstetrics: The Official Organ of the International Federation of Gynaecology and Obstetrics*. 2009 Dec;107(3):289–294. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19846089>.
- [120] Chaillet N, Dubé E, Dugas M, Audibert F, Tourigny C, Fraser WD, et al. Evidence-based strategies for implementing guidelines in obstetrics: a systematic review. *Obstetrics and Gynecology*. 2006 Nov;108(5):1234–1245. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17077251>.
- [121] Jamtvedt G, Young JM, Kristoffersen DT, O’Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews (Online)*. 2006;(2):CD000259. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16625533>.
- [122] Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: Making feedback actionable. *Implementation Science: IS*. 2006;1:9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16722539>.
- [123] Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. *Social Science & Medicine (1982)*. 2010 May;70(10):1618–1625. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20207464>.

- [124] Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Quality & Safety in Health Care*. 2005 Feb;14(1):26–33. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15692000>.
- [125] Larson EL, Patel SJ, Evans D, Saiman L. Feedback as a strategy to change behaviour: the devil is in the details. *Journal of Evaluation in Clinical Practice*. 2011; Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2753.2011.01801.x/abstract>.
- [126] Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*. 2008 May;27(3):379–387. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18624603>.
- [127] Carver CS, Scheier MF. Control theory: A useful conceptual framework for personality – social, clinical, and health psychology. *Psychological Bulletin* Vol 92(1). 1982 Jul;92(1):111–135.
- [128] Siddiqi K, Newell J, Robinson M. Getting evidence into practice: what works in developing countries? *International Journal for Quality in Health Care*. 2005 Oct;17(5):447–454. Available from: <http://intqhc.oxfordjournals.org/content/17/5/447.abstract>.
- [129] Pattinson RC, Say L, Makin JD, Bastos MH. Critical incident audit and feedback to improve perinatal and maternal mortality and morbidity. *Cochrane Database of Systematic Reviews* (Online). 2005;(4):CD002961. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16235307>.
- [130] van den Akker T, van Rhenen J, Mwagomba B, Lommerse K, Vinkhumbo S, van Roosmalen J. Reduction of severe acute maternal morbidity and maternal mortality in Thyolo District, Malawi: the impact of obstetric audit. *PloS One*. 2011;6(6):e20776. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21677788>.
- [131] Bakker W, van den Akker T, Mwagomba B, Khukulu R, van Elteren M, van Roosmalen J. Health workers' perceptions of obstetric critical incident audit in Thyolo District, Malawi. *Tropical Medicine & International Health: TM & IH*. 2011 Jul; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21767335>.
- [132] Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Family Practice*. 2000 Feb;17 Suppl 1:S11–16. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10735262>.
- [133] Grol RPTM, Bosch MC, Hulscher MEJL, Eccles MP, Wensing M. Planning and studying improvement in patient care: the use of theoretical perspectives. *The Milbank quarterly*. 2007;85(1):93–138.

- [134] Oxman AD, Fretheim A, Flottorp S. The OFF theory of research utilization. *Journal of clinical epidemiology*. 2005 Feb;58(2):113–116; discussion 117–120.
- [135] Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of clinical epidemiology*. 2005 Feb;58(2):107–112.
- [136] Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation science: IS*. 2012;7:37.
- [137] Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*. 1977;84(2):191–215.
- [138] Michie S, Stralen MMv, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*. 2011 Apr;6(1):42. Available from: <http://www.implementationscience.com/content/6/1/42/abstract>.
- [139] Vicente KJ. *Cognitive work analysis toward safe, productive, and healthy computer-based work*. Mahwah, N.J.: Lawrence Erlbaum Associates; 1999.
- [140] Ilgen DR, Fisher CD, Taylor MS. Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*. 1979;64(4):349–371.
- [141] Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education*. 2012 Mar;17(1):15–26. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3274671/>.
- [142] Parkes J, Abercrombie S, McCarty T. Feedback sandwiches affect perceptions but not performance. *Advances in health sciences education: theory and practice*. 2013 Aug;18(3):397–407.
- [143] Judd C. Practice without knowledge of results. *Psychological Monographs*. 1905 Mar;7(1):185–198.
- [144] Baker DF, Buckley MR. A historical perspective of the impact of feedback on behaviour. *Journal of Management History (Archive)*. 1996 Jan;2(4):21–33. Available from: <http://www.emeraldinsight.com/journals.htm?articleid=871900&show=html>.
- [145] Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* Vol 57(9). 2002 Sep;57(9):705–717.

- [146] Parrino TA. The nonvalue of retrospective peer comparison feedback in containing hospital antibiotic costs. *The American Journal of Medicine*. 1989 Apr;86(4):442–448. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2929630>.
- [147] Kinicki AJ, Prussia GE, Wu BJ, McKee-Ryan FM. A covariance structure analysis of employees' response to performance feedback. *The Journal of Applied Psychology*. 2004 Dec;89(6):1057–1069. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15584841>.
- [148] Dijk DVAN, Kluger AN. Task type as a moderator of positive / negative feedback effects on motivation and performance : A regulatory focus perspective. *Library*. 2010;32(July).
- [149] Kluger AN, DeNisi A. Feedback Interventions: Toward the Understanding of a Double-edged Sword. *Current Directions in Psychological Science*. 1998 Jun;7(3):67–72. Available from: <http://cdp.sagepub.com/content/7/3/67.short>.
- [150] Kluger AN, Van Dijk D. Feedback, the various tasks of the doctor, and the feedforward alternative. *Medical Education*. 2010 Dec;44(12):1166–1174. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21091758>.
- [151] Higgins ET. Beyond pleasure and pain. *American Psychologist*. 1997;52(12):1280–300.
- [152] Higgins ET, Shah J, Friedman R. Emotional responses to goal attainment: strength of regulatory focus as moderator. *Journal of Personality and Social Psychology*. 1997;72(3):515–525.
- [153] Van Dijk D, Kluger AN. Feedback Sign Effect on Motivation: Is it Moderated by Regulatory Focus? *Applied Psychology*. 2004 Jan;53(1):113–135. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1464-0597.2004.00163.x/abstract>.
- [154] Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Understanding responses to feedback: the potential and limitations of regulatory focus theory. *Medical education*. 2012 Jun;46(6):593–603. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22626051>.
- [155] Pritchard RD, Jones SD, Roth PL, Stuebing KK, Ekeberg SE. Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology*. 1988;73(2):337–358.
- [156] DeShon RP, Kozlowski SWJ, Schmidt AM, Milner KR, Wiechmann D. A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *The Journal of Applied Psychology*. 2004 Dec;89(6):1035–1056. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15584840>.
- [157] Nadler DA. The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance*. 1979

- Jun;23(3):309–338. Available from: <http://www.sciencedirect.com/science/article/pii/0030507379900011>.
- [158] Hinsz VB, Tindale RS, Vollrath DA. The emerging conceptualization of groups as information processors. *Psychological Bulletin*. 1997 Jan;121(1):43–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9000891>.
 - [159] Van der Vegt GS, de Jong SB, Bunderson JS, Molleman E. Power Asymmetry and Learning in Teams: The Moderating Role of Performance Feedback. *Organization Science*. 2010 Mar;21(2):347–361. Available from: <http://orgsci.journal.informs.org/content/21/2/347.abstract>.
 - [160] Kreuter MW, Skinner CS. Tailoring: what’s in a name? *Health Education Research*. 2000 Feb;15(1):1–4. Available from: <http://her.oxfordjournals.org/content/15/1/1>.
 - [161] Baker R, Camosso-Stefinovic J, Gillies C, Shaw EJ, Cheater F, Flottorp S, et al. Tailored interventions to overcome identified barriers to change: effects on professional practice and health care outcomes. *The Cochrane database of systematic reviews*. 2010;(3):CD005470.
 - [162] Heather Colquhoun, Grimshaw, Jeremy, Wensing, Michel. Chapter 3.3b Mapping KT interventions to barriers and facilitators. In: Straus S, Tetroe J, Graham ID, editors. *Knowledge Translation in Health Care: Moving from Evidence to Practice*. John Wiley & Sons; 2013. .
 - [163] *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. 3rd ed. New York: Springer; 2006.
 - [164] Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA: the journal of the American Medical Association*. 2005 Mar;293(10):1223–1238.
 - [165] Crowley RS, Medvedeva O. An intelligent tutoring system for visual classification problem solving. *Artificial intelligence in medicine*. 2006 Jan;36(1):85–117.
 - [166] Peleg M. Computer-interpretable clinical guidelines: a methodological review. *Journal of biomedical informatics*. 2013 Aug;46(4):744–763.
 - [167] Schulz S, Jansen L. Formal ontologies in biomedical knowledge representation. *Year-book of Medical Informatics*. 2013;8(1):132–146.
 - [168] Gruber TR. A translation approach to portable ontology specifications. *Knowledge Acquisition*. 1993 Jun;5(2):199–220. Available from: <http://www.sciencedirect.com/science/article/pii/S1042814383710083>.

- [169] Brewster C, O'Hara K. Knowledge representation with ontologies: the present and future. *IEEE Intelligent Systems*. 2004 Jan;19(1):72–81.
- [170] Rosse C, Mejino JLV. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*. 2003 Dec;36(6):478–500.
- [171] Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. *Yearbook of Medical Informatics*. 2006;p. 124–135.
- [172] Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association*. 2012 Sep;p. amiajnl-2012-001145. Available from: <http://jamia.bmj.com/content/early/2012/09/30/amiajnl-2012-001145>.
- [173] Conati C, Maclaren H. Data-Driven Refinement of a Probabilistic Model of User Affect. In: *Proceedings of the 10th International Conference on User Modeling*. UM'05. Berlin, Heidelberg: Springer-Verlag; 2005. p. 40–49. Available from: http://dx.doi.org/10.1007/11527886_7.
- [174] McQuiggan SW, Mott BW, Lester JC. Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-Adapted Interaction*. 2008 Feb;18(1-2):81–123. Available from: <http://link.springer.com/article/10.1007/s11257-007-9040-y>.
- [175] Horvitz E, Kadie C, Paek T, Hovel D. Models of Attention in Computing and Communication: From Principles to Applications. *Commun ACM*. 2003 Mar;46(3):52–59. Available from: <http://doi.acm.org/10.1145/636772.636798>.
- [176] Parrott R. Emphasizing “Communication” in Health Communication. *Journal of Communication*. 2004 Dec;54(4):751–787. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1460-2466.2004.tb02653.x/abstract>.
- [177] Kreuter M. *Tailoring health messages customizing communication with computer technology*. Mahwah, N.J.: L. Erlbaum; 2000.
- [178] Park EJ, McDaniel A, Jung MS. Computerized tailoring of health information. *Computers, informatics, nursing: CIN*. 2009 Feb;27(1):34–43.
- [179] Kreuter MW, Wray RJ. Tailored and Targeted Health Communication: Strategies for Enhancing Information Relevance. *American Journal of Health Behavior*. 2003 Nov;27(1):S227–S232.
- [180] Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. *Patient education and counseling*. 1999 Feb;36(2):107–117.
- [181] Marlin B, Adams R, Sadasivam R, Houston T. Towards Collaborative Filtering Recommender Systems for Tailored Health Communications. *Proceedings of the AMIA Symposium*. 2013 Nov;p. 1600–1607.

- [182] Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*. 2003 Jan;7(1):76–80.
- [183] Herlocker JL, Konstan JA, Borchers A, Riedl J. An Algorithmic Framework for Performing Collaborative Filtering. In: *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '99. New York, NY, USA: ACM; 1999. p. 230–237. Available from: <http://doi.acm.org/10.1145/312624.312682>.
- [184] Krebs P, Prochaska JO, Rossi JS. DEFINING WHAT WORKS IN TAILORING: A META-ANALYSIS OF COMPUTERTAILORED INTERVENTIONS FOR HEALTH BEHAVIOR CHANGE. *Preventive medicine*. 2010;51(3-4):214–221. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2939185/>.
- [185] Landis Lewis Z, Mello-Thoms C, Gadabu OJ, Gillespie EM, Douglas GP, Crowley RS. The feasibility of automating audit and feedback for ART guideline adherence in Malawi. *Journal of the American Medical Informatics Association: JAMIA*. 2011 Sep; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21565857>.
- [186] Shiffman RN, Karras BT, Agrawal A, Chen R, Marengo L, Nath S. GEM: a proposal for a more comprehensive guideline document model using XML. *Journal of the American Medical Informatics Association: JAMIA*. 2000 Oct;7(5):488–498. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10984468>.
- [187] Zachariah R, Harries K, Moses M, Manzi M, Line A, Mwagomba B, et al. Very early mortality in patients starting antiretroviral treatment at primary health centres in rural Malawi. *Tropical Medicine & International Health*. 2009 Jul;14(7):713–721. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3156.2009.02291.x/abstract>.
- [188] Beadles WI, Jahn A, Weigel R, Clutterbuck D. Peripheral neuropathy in HIV-positive patients at an antiretroviral clinic in Lilongwe, Malawi. *Tropical Doctor*. 2009 Apr;39(2):78–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19299286>.
- [189] Lowrance D, Makombe S, Harries A, Yu J, Aberle-Grasse J, Eiger O, et al. Lower early mortality rates among patients receiving antiretroviral treatment at clinics offering cotrimoxazole prophylaxis in Malawi. *Journal of Acquired Immune Deficiency Syndromes (1999)*. 2007 Sep;46(1):56–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17972365>.
- [190] Crabtree BF, Miller WL, editors. *Doing Qualitative Research*. Sage Publications, Inc; 1992.
- [191] Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. *MIS Q*. 2003 Sep;27(3):425–478. Available from: <http://dl.acm.org/citation.cfm?id=2017197.2017202>.

- [192] Friedland G, Harries A, Coetzee D. Implementation Issues in Tuberculosis/HIV Program Collaboration and Integration: 3 Case Studies. *The Journal of Infectious Diseases*. 2007 Jul;196(s1):S114–S123. Available from: http://jid.oxfordjournals.org/content/196/Supplement_1/S114.short.
- [193] Ong MS, Magrabi F, Coiera E. Syndromic surveillance for health information system failures: a feasibility study. *Journal of the American Medical Informatics Association: JAMIA*. 2013 May;20(3):506–512.
- [194] HEALTHQUAL Performance Measurement Report. HEALTHQUAL International; 2013. Available from: <http://www.healthqual.org/healthqual-performance-measurement-report>.
- [195] Anseel F, Beatty AS, Shen W, Lievens F, Sackett PR. How Are We Doing After 30 Years? A Meta-Analytic Review of the Antecedents and Outcomes of Feedback-Seeking Behavior. *Journal of Management*. 2013 May; Available from: <http://jom.sagepub.com/content/early/2013/05/09/0149206313484521>.
- [196] Silver WS, Mitchell TR, Gist ME. Responses to Successful and Unsuccessful Performance: The Moderating Effect of Self-Efficacy on the Relationship between Performance and Attributions. *Organizational Behavior and Human Decision Processes*. 1995 Jun;62(3):286–299. Available from: <https://sremote.pitt.edu/science/article/pii/S0749597885710515>.
- [197] Bandura A. Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist*. 1993;28(2):117–148. Available from: http://www.tandfonline.com/doi/abs/10.1207/s15326985ep2802_3.
- [198] Bandura A, Jourden FJ. Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology*. 1991;60(6):941–951.
- [199] Colquhoun H, Leeman J, Michie S, Lokker C, Bragge P, Hempel S, et al. Towards a common terminology: a simplified framework of interventions to promote and integrate evidence into health practices, systems, and policies. *Implementation Science*. 2014 May;9(1):51. Available from: <http://www.implementationscience.com/content/9/1/51/abstract>.
- [200] Waltz TJ, Powell BJ, Chinman MJ, Smith JL, Matthieu MM, Proctor EK, et al. Expert recommendations for implementing change (ERIC): protocol for a mixed methods study. *Implementation Science*. 2014 Mar;9(1):39. Available from: <http://www.implementationscience.com/content/9/1/39/abstract>.
- [201] Dehmer GJ, Drozda JP Jr, Brindis RG, Masoudi FA, Rumsfeld JS, Slattery LE, et al. Public Reporting of Clinical Quality Data: An Update for Cardiovascular Specialists. *Journal of the American College of Cardiology*. 2014 Apr;63(13):1239–1245.

- [202] Trautner BW, Petersen NJ, Hysong SJ, Horwitz D, Kelly PA, Naik AD. Overtreatment of asymptomatic bacteriuria: Identifying provider barriers to evidence-based care. *American journal of infection control*. 2014 Apr;.
- [203] Ryan C, Ross S, Davey P, Duncan EM, Fielding S, Francis JJ, et al. Junior doctors' perceptions of their self-efficacy in prescribing, their prescribing errors and the possible causes of errors. *British Journal of Clinical Pharmacology*. 2013 Dec;76(6):980–987. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/bcp.12154/abstract>.
- [204] Eccles MP, Johnston M, Hrisos S, Francis J, Grimshaw J, Steen N, et al. Translating clinicians' beliefs into implementation interventions (TRACII): A protocol for an intervention modeling experiment to change clinicians' intentions to implement evidence-based practice. *Implementation science : IS*. 2007 Aug;2:27. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1988805/>.
- [205] Walker AE, Grimshaw JM, Armstrong EM. Salient beliefs and intentions to prescribe antibiotics for patients with a sore throat. *British journal of health psychology*. 2001 Nov;6(Part 4):347–360.
- [206] Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine*. 2013 Mar;.
- [207] Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ (Clinical research ed)*. 2014;348:g3725.
- [208] Mislevy RJ, Gitomer DH. The role of probability-based inference in an intelligent tutoring system. *User Modeling and User-Adapted Interaction*. 1995 Sep;5(3-4):253–282. Available from: <http://link.springer.com/article/10.1007/BF01126112>.
- [209] Braspenning J, Hermens R, Calsbeek H, Westert G, Campbell S, Grol R. Quality and safety of care: the role of indicators. In: Grol R, Wensing M, Eccles r, Davis D, editors. *Improving Patient Care*. John Wiley & Sons, Ltd; 2013. p. 115–136. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/9781118525975.ch7/summary>.